

GenCore Version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2003, 08:53:46 ; Search time 67 Seconds

(Without alignments)
1020.262 Million cell updates/sec

Title: US-09-502-426a-2

Perfect score: 2681

Sequence: 1 MEETENHTLLPLLLPSLLS.....FAFPYDFPGLPIRVSRL 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Geneseq_101002.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681	100.0	513	21	AA807921 A cytochrome P450
2	978.5	36.5	472	18	AAW27153 Arabidopsis thalia
3	978.5	36.5	472	21	AAW27153 Arabidopsis thalia
4	978.5	36.5	472	21	AAW27153 Arabidopsis thalia
5	978.5	36.5	472	21	AAW27153 Arabidopsis thalia
6	978.5	36.5	472	21	AAW27153 Arabidopsis thalia
7	978.5	36.5	472	21	AAW27153 Arabidopsis thalia
8	978.5	36.5	472	21	AAW27153 Arabidopsis thalia
9	978.5	36.5	472	21	AAW27153 Arabidopsis thalia
10	978.5	36.5	472	21	AAW27153 Arabidopsis thalia

11	726	27.1	461	21	AAW27153 Arabidopsis thalia
12	726	27.1	461	21	AAW27153 Arabidopsis thalia
13	726	27.1	461	21	AAW27153 Arabidopsis thalia
14	726	27.1	461	21	AAW27153 Arabidopsis thalia
15	726	27.1	461	21	AAW27153 Arabidopsis thalia
16	726	27.1	461	21	AAW27153 Arabidopsis thalia
17	726	27.1	461	21	AAW27153 Arabidopsis thalia
18	726	27.1	461	21	AAW27153 Arabidopsis thalia
19	726	27.1	461	21	AAW27153 Arabidopsis thalia
20	726	27.1	461	21	AAW27153 Arabidopsis thalia
21	726	27.1	461	21	AAW27153 Arabidopsis thalia
22	726	27.1	461	21	AAW27153 Arabidopsis thalia
23	726	27.1	461	21	AAW27153 Arabidopsis thalia
24	726	27.1	461	21	AAW27153 Arabidopsis thalia
25	726	27.1	461	21	AAW27153 Arabidopsis thalia
26	726	27.1	461	21	AAW27153 Arabidopsis thalia
27	726	27.1	461	21	AAW27153 Arabidopsis thalia
28	726	27.1	461	21	AAW27153 Arabidopsis thalia
29	726	27.1	461	21	AAW27153 Arabidopsis thalia
30	726	27.1	461	21	AAW27153 Arabidopsis thalia
31	726	27.1	461	21	AAW27153 Arabidopsis thalia
32	726	27.1	461	21	AAW27153 Arabidopsis thalia
33	726	27.1	461	21	AAW27153 Arabidopsis thalia
34	726	27.1	461	21	AAW27153 Arabidopsis thalia
35	726	27.1	461	21	AAW27153 Arabidopsis thalia
36	726	27.1	461	21	AAW27153 Arabidopsis thalia
37	726	27.1	461	21	AAW27153 Arabidopsis thalia
38	726	27.1	461	21	AAW27153 Arabidopsis thalia
39	726	27.1	461	21	AAW27153 Arabidopsis thalia
40	726	27.1	461	21	AAW27153 Arabidopsis thalia
41	726	27.1	461	21	AAW27153 Arabidopsis thalia
42	726	27.1	461	21	AAW27153 Arabidopsis thalia
43	726	27.1	461	21	AAW27153 Arabidopsis thalia
44	726	27.1	461	21	AAW27153 Arabidopsis thalia
45	726	27.1	461	21	AAW27153 Arabidopsis thalia

ALIGNMENTS

RESULT 1	AA807921	standard; Protein: 513 AA.
XX	AA807921:	
AC	14-NOV-2000	(first entry)
DT	14-NOV-2000	(first entry)
XX		
DE	A cytochrome P450 enzyme designated DMF4.	
KW	DMF4: cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;	
KW	plant phenotype; cell elongation.	
XX		
OS	Arabidopsis sp.	
PN	W0200047715-A2.	
XX		
PD	17-AUG-2000.	
XX		
XX	11-FEB-2000; 2000WO-US03820.	
PF	11-FEB-1999; 99US-0119657.	
PR	11-FEB-1999; 99US-0119658.	
XX		
PA	(ARIZ-) ARIZONA BOARD OF REGENTS.	
XX		
PI	Azpiroz R, Choe S, Feldmann KA;	
XX		
DR	WPI: 2000-549142/50.	
XX	N-PSDB: AAA59599.	
XX		
PT	New isolated dwf4 polynucleotide useful for altering the phenotype of plants, for diagnostic assays and in the production of antibodies -	

XX Claim 50; Fig 11; 113pp; English.
 XX
 CC The present sequence represents a DWF4 polypeptide. The polypeptide is a
 CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
 CC brassinosteroids. Specifically, it mediates multiple
 CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
 CC polynucleotide is used for altering the phenotype of a plant. DWF4
 CC plants display a dramatic reduction in the length of different organs,
 CC and this size reduction is attributable to a defect in cell elongation.
 CC The DWF4 polynucleotides and polypeptides can be used in diagnostic
 CC assays and to generate antibodies, which can be used to produce
 CC immunogenic compositions.
 CC
 XX
 XX Sequence 513 AA:

Query Match 100.0%; Score 2681; DB 21; Length 513;
 Best Local Similarity 100.0%; Pred. No. 1.3e-228;
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METEHTHTLLPLLLPSLLSLLEFLILKRRNRKTRFNPPEKSGMPFAGETIGYAKPYT 60
 DB 1 METEHTHTLLPLLLPSLLSLLEFLILKRRNRKTRFNPPEKSGMPFAGETIGYAKPYT 60
 QY 61 ATTLDGFMQOHVSKYKIRSNLFGSEPTVSADAGLNREFTLQNEGRLEFCSYPRSTIGIL 120
 DB 61 ATTLDGFMQOHVSKYKIRSNLFGSEPTVSADAGLNREFTLQNEGRLEFCSYPRSTIGIL 120
 QY 121 GKMSMLVGVGDHNRDMRSISLNFSLSHARLRTILKDVENHTLVLDMSQONSIFSQDRA 180
 DB 121 GKMSMLVGVGDHNRDMRSISLNFSLSHARLRTILKDVENHTLVLDMSQONSIFSQDRA 180
 QY 121 GKMSMLVGVGDHNRDMRSISLNFSLSHARLRTILKDVENHTLVLDMSQONSIFSQDRA 180
 DB 121 GKMSMLVGVGDHNRDMRSISLNFSLSHARLRTILKDVENHTLVLDMSQONSIFSQDRA 180
 QY 181 KKFETFLMAKHIMSMDPGEETEQLKKEYVTYPMKGVASPLNPGTAYHKAQSRATITIK 240
 DB 181 KKFETFLMAKHIMSMDPGEETEQLKKEYVTYPMKGVASPLNPGTAYHKAQSRATITIK 240
 QY 241 FIERKKEERKLDIKEDQEEVEKYTDEDAEMSKSDHVRKORTDDDLGWLKHSNLSSTQ 300
 DB 241 FIERKKEERKLDIKEDQEEVEKYTDEDAEMSKSDHVRKORTDDDLGWLKHSNLSSTQ 300
 QY 301 IIDLILSLFAGHETSSVALAIFFLQACPAVEELREHLEIARAKKEGSELNMD 360
 DB 301 IIDLILSLFAGHETSSVALAIFFLQACPAVEELREHLEIARAKKEGSELNMD 360
 QY 361 YKKMOTQOVINETLRLGVNVRFLHRAKALDVRKYGDIPSGMKVLPVISAHLDSRYD 420
 DB 361 YKKMOTQOVINETLRLGVNVRFLHRAKALDVRKYGDIPSGMKVLPVISAHLDSRYD 420
 QY 421 QPNLFPNPMQOONNGASSSGSGSFSTWGNNTMPFGGGRCLCAGSELAKLEMAVFTHILY 480
 DB 421 QPNLFPNPMQOONNGASSSGSGSFSTWGNNTMPFGGGRCLCAGSELAKLEMAVFTHILY 480
 QY 481 LKFNWELADDDQPAFPVDFPNGLPIRVSRIL 513
 DB 481 LKFNWELADDDQPAFPVDFPNGLPIRVSRIL 513

RESULT 2
 AAW27153
 ID AAW27153 standard; Protein; 472 AA.

XX AAW27153;
 AC
 XX
 XX 14-APR-1998 (first entry)
 DT
 XX Arabidopsis thaliana cytochrome P450-type hydroxylase.
 DE
 XX
 XX Cytochrome P450-type hydroxylase; identification; brassinosteroid;
 KW brassinosteroid inhibitor; modified plant; recombinant production;
 KW teasterone.
 XX
 OS Arabidopsis thaliana.
 XX

PN W09735986-A1.
 XX
 XX 02-OCT-1997.
 PD
 XX
 XX 27-MAR-1997; 97WO-EP01586.
 PF
 XX
 XX 27-MAR-1996; 96US-0622166.
 PR
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA Allmann T, Koncz C, Mathur J, Szekeres MA;
 DR WPI: 1997-489649/45.
 DR N-PSDB: AAT85306, AAT85307.
 DR
 XX
 XX New isolated plant cytochrome P450-type hydroxylase gene - used to
 PT identify substances acting as brassino-steroid(s) or brassinosteroid
 PT inhibitors for the production of modified plants
 PT
 XX
 XX Claim 1; Pages 44-46; 77pp; English.

CC The present sequence is Arabidopsis thaliana cytochrome
 CC P450-type hydroxylase. The hydroxylase can be used to identify
 CC brassinosteroids or brassinosteroid inhibitors, useful to produce
 CC plants with modified physiological and/or phenotypic
 CC characteristics. The modified plants may show, e.g. stimulation of
 CC growth, increased cell elongation, increased wood production,
 CC accelerated seed germination at low temperatures, an increase in
 CC dry weight, repressed anthocyanin production during growth in light
 CC and/or inhibited de-etiolation which is induced, e.g. by cytokinin,
 CC in the dark or an increase in stress tolerance. The hydroxylase or
 CC its coding sequence can also be used for the recombinant production
 CC of compounds, e.g. teasterone.
 CC
 XX

Sequence 472 AA:

Query Match 36.5%; Score 978.5; DB 18; Length 472;
 Best Local Similarity 41.9%; Pred. No. 6.6e-78;
 Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLLSLLEFLILKRRNRKTRFNPPEKSGMPFAGETIGYAKPYTATTLDGFMQOH 71
 DB 7 LLLSSIAAGPL--LLLRTRRRKGRPLPGSLGILPILGAYTENEPPIDR 63
 QY 72 VSKYGIYYSNLFGEPTIYSADAGLNREFTLQNEGRLEFCSYPRSTIGILGKMSMLVLD 131
 DB 64 VARGSVFMTLHGEPTISADPETNRFVLQNEGRLEFCSYPRSTIGILGKMSMLVLD 123
 QY 132 MHRDMRSISLNFSLSHARLRTILKDVENHTLVLDMSQONSIFSQDRAKKEFENLMAR 191
 DB 124 LKRNHSLTMSFRANSSTIIDHMLDIDRLVRRNLSWSSRYLL--MEKAKTTFELTVKQ 181
 QY 192 INSMQGEETEOLKKEYVTYPMKGVASPLNPGTAYHKAQSRATITIKFIERKMEERL 251
 DB 182 LMSFDPG-EMSSSLRKEVLLVIEGFPSPPLPSTYTYRAIQAR-----RKVAEALT 232
 QY 252 DIKEDEQEEVEKYTDEDAEMSKSDHVRKORTDDDLGWLKHSNLSSTQIIDLILSLFA 311
 DB 233 VVYMKRREEEEGAE-----RKDDMLAALLAADGSDDEIVDFVALLVA 278
 QY 312 GHETSSVALAIFFLQACPAVEELREHLEIARAKKEGSELNMDQYKKMDFQCVY 371
 DB 279 GYEITSTITLAVKFLTEPFLAOLKEHEKIRANKSD--SYLSEMSYKSNPTQCVV 336
 QY 372 NETLRLGVNVRFLHRAKALDVRKYGDIPSGMKVLPVISAHLDSRYDQPNLFPNWRQ 431
 DB 337 NETLRLVANIIGVFRRAMTDEIKGYKIPKGMKVSFFAVAHDPNHFADARFENWRQ 396
 QY 432 QONNGASSSGSFSTWGNNTMPFGGGRCLCAGSELAKLEMAVFTHILYLFNWLAD 490
 DB 397 -----SNSVTTGSPSNVETPFGGGRCLCAGSELAKLEMAVFTHILYLVTSFVPAQ 446
 QY 491 DQPAFPVDFPNGLPIRVSR 511

Db 447 DKLVFPTRTRQKRYPIFYKR 467

RESULT 3
ID AAG44571 standard; Protein; 472 AA.
AC AAG44571;
XX
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55847.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI03405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

```
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149903.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156599.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.
```

Query Match 36.5% Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9% Pred. No. 6.6e-76;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

```
QY 12 LLLPSTLSTLFLTLRRNRKTFENLPKSGWPFLEGTIGYLLKPYTATTLDGPMOQH 71
DB 7 LLLLSSTIAGFL---LLRRRTYRRMGILPPGSLGLPLGRTFOLLGAKTKTENPEPFIDR 63
QY 72 VSKYGIKIRSNLFGEPITVSADAGINRFTLQNEGRLFCSTYPRSTIGILGRKMSMLVIGD 131
DB 64 VARYGSVMTLFGPITFSADPETNRFLVQNEGKLFCSYPAASICNLGKHSLLMKGS 123
```

```
QY 133 MHRDRSISLNLFSARLFTLLKQVERHTLFLVDSMOONSIFSAODEAKKFTFNLMAKH 191
DB 124 LHKRHSLSLMSFANSIIIDHMLDRLVRRNLSSWSSRVL--MEAKKITFELTVQ 181
QY 192 IMSMDPGEETEQLKKEYTFEMKGVASAPNLPGTAVHKAALOSRATILKFIKKMEERL 251
DB 182 LMSFDPG-EWSLSLKRKYLVTGFFSLPLPSTTYRAIAR-----RKVAELT 232
QY 252 DIKEEDQEEVEVKTDEAEMSKSDVHRKQRTDDDLGWLKHSNLTSTEQIIDLILSLFA 311
DB 233 VVMMKRREEEEGAE-----RKMDMLAALLAADGSDDEIYDFVALLVA 278
QY 312 GHETSSVALAIAIFLQACPAAVEELREHLEIAKKELGSELMNDYKMDSTQCVY 371
DB 279 GYETSTITMLVAKFLETPPLAOLKEHEKIRAMKSD--SYSTEMSDYKSMPTQCVY 336
QY 372 NETLRGNVVRFLHKKALKDVRKGDIPSGKVLPIVSAVILDSRRDQPLFPMRMQ 431
DB 337 NETLRVANITGCVFRAMADVETKGTIKPKMKVSSFRVAHLDPNHFADARTFPMRMQ 396
QY 432 QONNGASSSGSGSFSTWGN--YMPGGGPRLCAGSELAKLENAVFIHLVLKFMELAE 490
DB 397 -----SNSVTGPGSNVFTPPGGGPRLCGYELARVALSVLHLRYTGFSVPRADQ 446
QY 491 DQFAFPVDFPNCGLPIRVSR 511
DB 447 DLVFEPTTRQKRYPFVKR 467
```

RESULT 4

AA045022 standard; Protein; 472 AA.

AA045022;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 56469.

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

Ep1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

```
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
```


PR 11-MAY-1999; 9905-0134256.
 PR 14-MAY-1999; 9905-0134218.
 PR 14-MAY-1999; 9905-0134219.
 PR 14-MAY-1999; 9905-0134221.
 PR 14-MAY-1999; 9905-0134320.
 PR 18-MAY-1999; 9905-0134768.
 PR 19-MAY-1999; 9905-0134941.
 PR 20-MAY-1999; 9905-0135124.
 PR 21-MAY-1999; 9905-0135353.
 PR 24-MAY-1999; 9905-0135629.
 PR 25-MAY-1999; 9905-0136021.
 PR 27-MAY-1999; 9905-0136392.
 PR 28-MAY-1999; 9905-0136782.
 PR 01-JUN-1999; 9905-0137222.
 PR 03-JUN-1999; 9905-0137528.
 PR 04-JUN-1999; 9905-0137502.
 PR 07-JUN-1999; 9905-0137724.
 PR 08-JUN-1999; 9905-0138094.
 PR 10-JUN-1999; 9905-0138540.
 PR 10-JUN-1999; 9905-0138847.
 PR 14-JUN-1999; 9905-0139119.
 PR 16-JUN-1999; 9905-0139452.
 PR 16-JUN-1999; 9905-0139453.
 PR 17-JUN-1999; 9905-0139492.
 PR 18-JUN-1999; 9905-0139454.
 PR 18-JUN-1999; 9905-0139455.
 PR 18-JUN-1999; 9905-0139456.
 PR 18-JUN-1999; 9905-0139457.
 PR 18-JUN-1999; 9905-0139458.
 PR 18-JUN-1999; 9905-0139459.
 PR 18-JUN-1999; 9905-0139460.
 PR 18-JUN-1999; 9905-0139461.
 PR 18-JUN-1999; 9905-0139462.
 PR 18-JUN-1999; 9905-0139463.
 PR 18-JUN-1999; 9905-0139750.
 PR 18-JUN-1999; 9905-0139763.
 PR 21-JUN-1999; 9905-0139817.
 PR 22-JUN-1999; 9905-0139899.
 PR 23-JUN-1999; 9905-0140353.
 PR 23-JUN-1999; 9905-0140354.
 PR 24-JUN-1999; 9905-0140695.
 PR 28-JUN-1999; 9905-0140823.
 PR 29-JUN-1999; 9905-0140991.
 PR 30-JUN-1999; 9905-0141287.
 PR 01-JUL-1999; 9905-0141442.
 PR 01-JUL-1999; 9905-0142154.
 PR 02-JUL-1999; 9905-0142055.
 PR 06-JUL-1999; 9905-0142390.
 PR 08-JUL-1999; 9905-0142803.
 PR 09-JUL-1999; 9905-0142920.
 PR 12-JUL-1999; 9905-0142977.
 PR 13-JUL-1999; 9905-0143542.
 PR 14-JUL-1999; 9905-0143624.
 PR 15-JUL-1999; 9905-0144005.
 PR 16-JUL-1999; 9905-0144085.
 PR 16-JUL-1999; 9905-0144086.
 PR 19-JUL-1999; 9905-0144325.
 PR 19-JUL-1999; 9905-0144331.
 PR 19-JUL-1999; 9905-0144332.
 PR 19-JUL-1999; 9905-0144333.
 PR 19-JUL-1999; 9905-0144334.
 PR 19-JUL-1999; 9905-0144335.
 PR 20-JUL-1999; 9905-0144352.
 PR 20-JUL-1999; 9905-0144632.
 PR 20-JUL-1999; 9905-0144684.
 PR 21-JUL-1999; 9905-0144814.
 PR 21-JUL-1999; 9905-0145086.
 PR 21-JUL-1999; 9905-0145088.
 PR 22-JUL-1999; 9905-0145085.
 PR 22-JUL-1999; 9905-0145087.
 PR 22-JUL-1999; 9905-0145089.
 PR 22-JUL-1999; 9905-0145192.
 PR 23-JUL-1999; 9905-0145145.

PR 23-JUL-1999; 9905-0145218.
 PR 23-JUL-1999; 9905-0145224.
 PR 26-JUL-1999; 9905-0145276.
 PR 27-JUL-1999; 9905-0145913.
 PR 27-JUL-1999; 9905-0145918.
 PR 27-JUL-1999; 9905-0145919.
 PR 28-JUL-1999; 9905-0145951.
 PR 02-AUG-1999; 9905-0146386.
 PR 02-AUG-1999; 9905-0146388.
 PR 02-AUG-1999; 9905-0146389.
 PR 03-AUG-1999; 9905-0147038.
 PR 04-AUG-1999; 9905-0147204.
 PR 04-AUG-1999; 9905-0147302.
 PR 05-AUG-1999; 9905-0147192.
 PR 05-AUG-1999; 9905-0147260.
 PR 06-AUG-1999; 9905-0147303.
 PR 06-AUG-1999; 9905-0147416.
 PR 09-AUG-1999; 9905-0147493.
 PR 09-AUG-1999; 9905-0147935.
 PR 10-AUG-1999; 9905-0148171.
 PR 11-AUG-1999; 9905-0148319.
 PR 12-AUG-1999; 9905-0148341.
 PR 13-AUG-1999; 9905-0148565.
 PR 13-AUG-1999; 9905-0148684.
 PR 16-AUG-1999; 9905-0149368.
 PR 17-AUG-1999; 9905-0149175.
 PR 18-AUG-1999; 9905-0149426.
 PR 20-AUG-1999; 9905-0149722.
 PR 20-AUG-1999; 9905-0149723.
 PR 20-AUG-1999; 9905-0149929.
 PR 23-AUG-1999; 9905-0149902.
 PR 23-AUG-1999; 9905-0149930.
 PR 25-AUG-1999; 9905-0150566.
 PR 26-AUG-1999; 9905-0150884.
 PR 27-AUG-1999; 9905-0151065.
 PR 27-AUG-1999; 9905-0151066.
 PR 27-AUG-1999; 9905-0151080.
 PR 30-AUG-1999; 9905-0151303.
 PR 31-AUG-1999; 9905-0151438.
 PR 01-SEP-1999; 9905-0151930.
 PR 07-SEP-1999; 9905-0152363.
 PR 10-SEP-1999; 9905-0153070.
 PR 13-SEP-1999; 9905-0153758.
 PR 15-SEP-1999; 9905-0154018.
 PR 16-SEP-1999; 9905-0154039.
 PR 20-SEP-1999; 9905-0154779.
 PR 22-SEP-1999; 9905-0155139.
 PR 23-SEP-1999; 9905-0155486.
 PR 24-SEP-1999; 9905-0155659.
 PR 28-SEP-1999; 9905-0156458.
 PR 29-SEP-1999; 9905-0156596.
 PR 04-OCT-1999; 9905-0157117.
 PR 05-OCT-1999; 9905-0157753.
 PR 06-OCT-1999; 9905-0157865.
 PR 07-OCT-1999; 9905-0158029.
 PR 08-OCT-1999; 9905-0158232.
 PR 12-OCT-1999; 9905-0158369.
 PR 13-OCT-1999; 9905-0158293.
 PR 13-OCT-1999; 9905-0159294.
 PR 13-OCT-1999; 9905-0159295.
 PR 14-OCT-1999; 9905-0159329.
 PR 14-OCT-1999; 9905-0159330.
 PR 14-OCT-1999; 9905-0159331.
 PR 14-OCT-1999; 9905-0159337.
 PR 14-OCT-1999; 9905-0159637.
 PR 14-OCT-1999; 9905-0159638.
 PR 18-OCT-1999; 9905-0159684.
 PR 21-OCT-1999; 9905-0160741.
 PR 21-OCT-1999; 9905-0160767.
 PR 21-OCT-1999; 9905-0160768.
 PR 21-OCT-1999; 9905-0160770.
 PR 21-OCT-1999; 9905-0160814.
 PR 21-OCT-1999; 9905-0160815.
 PR 22-OCT-1999; 9905-0160980.

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161923.
PR 29-OCT-1999; 99US-0162142.

Query Match 36.5%; Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9%; Pred. No. 6,6e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

OY 12 LLLLSLLSLFLLLKRRNRKTRFNLPKSGSMPFGETIGYKPYATILSGEMOQH 71
DB 7 LLLSSIAAGFL--LLKRTYRRNGLPSPGLPGLGETYQLIGAVKTENPEPPIDR 63
OY 72 VSKYKIVRSNLFGEPTIYSADAGINRFLQNEGRLEFECSPRSIGILGKSMVLVGD 131
DB 64 VARYGSVFETHLFGEPITFSADPEINRFVLQNEKLEFECSPASICNLGKSHLLMGKS 123
OY 132 MHRDRSISLNFSLHARLFTILKQVERHTLVLDSSWOONSIFSAODEAKKFTFNIMAKH 191
DB 134 LHKRHSLSLMSFANSIILKDHMLDIDRLVRFNLDSSWSRVLL--MEEAKKITFELTVQ 181
OY 192 IMSMDGEEETQLKKEVYTFMKGVVAPLNLGTAHYKALQSRATILKFIKKMEERL 251
DB 182 LMSFDPG-FWSSSLKREILVTEGFSPLPLFSTTYKKAIDAR-----KVAEALT 232
OY 232 DIKEBDOEEEVKTEDEAEMSKSDHVRKQRTDDLLGVYKHSNLSSTQILDLLIILFPA 311
DB 233 VYVMKREEEEGAE-----RKDKMLAALLAADGSDDEIVFVALLVA 278
OY 312 GHETSSVAIALAIFLQACPKAVEELREHLEIARAKKELGSELMDDYKMDTQCVI 371
DB 279 GETTSTMTLAVKFLTEPLALADLKEHEKIRAMKSD--SYLSWMSYKSMPTQCV 336
OY 372 NETLGNVVRFLHRRALKDVKYKGYDIPSGMKVLPVISAVALDMSRYQPLFPMRWQ 431
DB 337 NETLVANLIGVFRRAMTDEIKGKIPKGMKVSFRFAVLDHDPDARTFPMRWQ 396
OY 432 QONNGASSSGSFSTWGN--YMPFGGPRLCAGSELAKLEMAVFIHLVLKFNNELEAD 490
DB 397 -----SNSVTGFSNVFTPPGGPRLCGYELARVALSVLHRLVTFGFSVPAEQ 446
OY 491 DQFAPFPVDFPNGLPIRYSR 511
DB 447 DKLVEFPPTRTQKRYPIFKR 467

RESULT 5
AAG45021
ID AAG45021 standard; Protein; 491 AA.
XX AAG45021;
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56468.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-012180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

QY 491 DQFAFPYDFPENGPIRYSR 511
 DB 466 DKLVFPPTTRQKRYPIFYKR 486
 RESULT 6
 AAG44570
 ID AAG44570 standard; Protein; 492 AA.
 AAG44570;
 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 55846.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 OS Arabidopsis thaliana.
 PN EPI033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134768.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144000.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144864.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0146864.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 36.5%; Score 978.5; DB 21; Length 492;

Best Local Similarity 41.9%; Pred. No. 76-78; Mismatches 171; Indels 41; Gaps 8;

Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPILSLILLKRRNRKTRFNLPPKSGWPLGFTIGYLKPYATATTTGDFMQOH 71
DB 27 LLLISSIAAGFL--LLRRTYRRMGLPPGSLGLPLIGFTFOLIAKTEENPEPFI 83
QY 72 VSKYGRKRYRNLGCEPTIVASADAGLNRTIQNGRGLFECSPYPSIGITLKMMLVYGD 131

DB 84 VARYGVFWTHLGEPTTFISADPTNRFVLQNEGKLFECSPASICNLGKHSLLMKGS 143
QY 132 MHRDRSISLNLFLSHARLRTILTKDVERHTFLVDSMOONSIFSAODEAKFFENLMAKH 191
DB 144 LHKRMSLJMSFANSSIIKOHIMLIDIRLVFNLDSSRVL--MERAKITELVYKQ 201
QY 192 IMSMDPEEETEOLKKEYVFMKGVASAPNLPGTAYHAKLQSRATILKFERMEERKL 251
DB 202 LMSFDPG-EMSESLRKKEYLVIEGFSLPLPLFTYTRKALQAR-----RKVAELT 252
QY 252 DIKEDEDEBEERKTEDAEKSKSDHVKORDDLLGWLKHSULSTEOILLLSLFLA 311
DB 253 VVVKRRREEEGAE-----RKDKMLALLAADDFSDOEIYDFVALLVA 298
QY 312 GHETSSVAIALAIFLOACPAKVELEBEHEIARAKEGESEBLNDDYKKMFTOCVI 371
DB 299 GYETTSITMLTAVKFLIETPLALNQLKEHEKTIAMKSD--SYLEMSDYKSMFTQCIV 356
QY 372 NETLRGNVVFLLHKKALKDVRKGYDIPSGKYLPIYSAVHLNDSRYDQNLFPNRMQ 431
DB 357 NETLRVANIIGVFRRAMTDVEIKGYKIPKGMKVFSSPRAVHLDPNFKDARTFNPMRMQ 416
QY 432 QQNNGASSSGSSTWGN--YMPFGGPRLCASSELAKLEMAVFIHHLVLFKFWELAED 490
DB 417 -----SNSVTGPSNVTPEPGGPRLCPEYELARVALSVFLHRLVYGFQVPAEQ 466
QY 491 DQPPAPFVDPNGLPIRVS 511
DB 467 DKLVFFPTTRKRPYFVKR 487

RESULT 7
AAG44572
ID AAG44572 standard; Protein; 444 AA.

AC AAG44572;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55848.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PE 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 03-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

[illegible]


```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

```

```

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 29-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.18; Score 726; DB 21; Length 461;
Best Local Similarity 32.34; Pred. No. 1.5e-55;
Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

```

```

QY 12 LLLPSLLSLFLILKRRN--RKRPNLPKSGSPFLGRTIGYKLPYATITLGPMD 69
DB 7 LLIVSLCSALL-----RWNQMYTKNGLPPTGMPITGRTYFKQGP----NMR 55
QY 70 QHVSRYGKIYRSNLFGEPTIYSADAGLNRTIONEGRLEFCSYPSRISGILGKMSMLVLY 129
DB 56 NQRLRYGSPFKSHLGGPTILSMDSYVNRITLKESGIVPGYSQMLDILGTGNMAVH 115
QY 130 GDMHRDRSISLNFSLHARLFTILKDYERHTLFLDSMOONSIFSADEAKKFTF--- 185
DB 116 GSHRLARGSLISLSTWMDHILPKVDHMRSYLDQNNELFVIDIDOKKHAFLSSL 175
QY 186 -----NMAKHIMMDGGEETBOLKEEYTFMGVVSAPRLNCTGYAHKALQSRATLK 240
DB 176 TQIAGNLKRPV-----EEFKTAPFKLVVGTLSVPIDLTGTVRCGIOARNRND 225
QY 241 FIERKMEERKLDIKEEDQEEBEVTEDEAEKSKSDHYRKQRTDDDLGMYVKNHN---L 296
DB 226 LIRELMERR-----DSGE-----TFMDMGLYMKKKGKGNRYPL 258
QY 297 STEQILDILISLFAGHETSSVALAIAIFLOACPRAVEELREELHETARKEELGESET 356
DB 259 TDEIRIQVYVITISGVEYVSTSMALKYLDHPKALQELRABHLAFREKKRQ--DEPL 316
QY 357 NMWDYKKMDFTQCYINFTLRGNVRFHLRKALADVKYKGDIPSGKVLPIVSAVHLDN 416
DB 317 GLEDVKSMTFTRAVITYETSLRATVNGVLRKRTTDLLEINGLIRKWRITYVYTRREIN 376
QY 417 SRYPQPLFNPWRMOONNGASSSGSFSFTWGNMYMPFGGPRICAGSELAKLEMAVFI 476
DB 377 NLVEDPLIFNPWRMKMSLESG-----NSCYVFGGGLTLCRKEGLGIVEISFL 425
QY 477 HHVLKFNWELAEDDQPAPEVDFPGLPIRVS 510
DB 426 HYFVTRYRWEIEGDDELAVFPVPAPKGFHLRIS 459

RESULT 12
AAG46489
ID AAG46489 standard; Protein; 462 AA.
AC AAG46489;
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58493.
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

```

OS Arabidopsis thaliana.
XX
PM EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0121180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0136629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145146.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148177.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150567.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153756.

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.1%; Score 726; DB 21; Length 462;

Best Local Similarity 32.3%; Pred. No. 1.5e-55; Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

QY 12 LLLPPLSLLLFLILIKRN--RKTFRNLPFGKSGMPFGTIGLYKPTATTIGDFMQ 69
DB 8 LLIIVSLCSALL-----RNNQMKRYTKNGLPFGTGMGPIFGETTEFLKQGP----NFM 56
QY 70 OHVSKYKIRSNLFGPEPTVSADAGLNRFILNEGRLFGCSYPRISIGILGKWSMVLV 129
DB 57 NQRLRGSEFKSHLLGCPILISMDSEVNRKILKNESGLVPGYQSMIDLIGTCNMMAVH 116
QY 130 GDMHRDMRSISLNFSLHARLRTILKDVHRHTFLVLDSPQONSIFSAQDEAKKFTF--- 185
DB 117 GSSHRLMRGSLISLISSTMARDHLIPKVDHFMRSYLDQWHELEVIDIDQKTKHMAFSL 176
QY 186 -----NLMAHISMDBGEETEDOLKKEYVTPEMKGVASAPLNLPGYIYHKALOSRAITLK 240
DB 177 TQIAGNLRKPFV-----EEFKTAFKFLVGLTSPIDLPCTMYRGCIQARNHIDR 226
QY 241 FIERKMERLIDKKEEDQEEVEKTEDEAMSKSDHYRKQRTDDDLIGVYLKHSN-----L 296
DB 227 LILRELMOERR-----DSGE-----TTDDMLGYLKKKEGNRYPL 259
QY 297 STFOILDLILSLFAGHETSSVAIALAIFLQACPRAVEELREHETIARAKKEIGSEEL 356
DB 260 TDEIRIQVVTIILSGEYEVSTSMALKYLHDPKALQELRAHHLAFREKRKG--DEPL 317
QY 357 NMDDYKMDTQCVINETLGLGNVVRFLHRAKALKDVRYKGYDIPSGKVLDPVISAHLDN 416

DB 318 GLEDVSKMKFTRAVIYTSRLATLVNGVLKRTTDLINGILLPKGRIRIYYREINYDA 377
QY 417 SRYDQNLIFNPMWQOONNASSSGSFSSTGNNYMPFGGPRLCAGSELAKLEMAVFI 476
DB 378 NLKEDPLIFNPMWKMKSLSQ-----NSCFVFGGTRLCGKELGIVEISSFL 426
QY 477 HHVLKFNWELAEDDQPPAFPEYDFENGLPIRVS 510
DB 427 HYFVTRWMEIGDELVEPVPAPKGFHLRIS 460
RESULT 13
AAG46491
ID AAG46491 standard; Protein; 465 AA.
AC AAG46491;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58495.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
PD
XX 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149179.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151302.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155589.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.1%; Score 726; DB 21; length 465;
Best Local Similarity 32.3%; Pred. No. 1.5e-55;
Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;
QY 12 LLLPSSLLEFLILKRRN--RKRTRNLPFGKSGMFLGERTIGYLPYATTTIGDPMQ 69

Db 11 LITIVLSCALL-----RKNORRYTKNGLPETMGWPIFGETTERELKQCP-----NEMR 59
 QY 70 OHVSKYGIKYRNSLFEPTIVSDAGLNREILONEGRLEFCSPYRSIGGILGKWSMLVLY 129
 Db 60 NQRLRGSPFKSHLGCPTLISMDSEVNRILNENSGVLPGVPGSLDLGLGNCMAAVH 119
 QY 130 GDMHRRMRSISLNFSLHAKRLTLKADVERHTLFLVDSWOQNSIFSAODEAKFTF--- 185
 Db 120 GSSHRLMRGSLISLISSTMARHILPKVDHFMRSYLDQNNELVYIDQDKTKHMAFSL 179
 QY 186 -----NLMKHMISMDPGEETELKKEYTFPKGVASAPLNPGRYHMAISRAITLK 240
 Db 180 TQIAGNLRKPFV-----EEFTAFPKLVGTLSPIDLPNTYRCGIGQARNIDR 229
 QY 241 FIERKKEERKLDIKEEDQEEVEKTEDEAMSKSDHVRKQRTDDDLGWLKHSN---L 296
 Db 230 LRELMQERR-----DSGR-----TTMDLGYLMKKEGNRYPL 262
 QY 297 STEOILDLISLIFAGHETSSVAIALAIFPLQACPAVEELREHLEIARAKKELGESL 356
 Db 263 TDEIRDOVYTLISGSEYEVSTSMALKYLDHPKALQELRAEHILAFRRKRO--DEPL 320
 QY 357 NMDDYKMKPOTQCVINETLRLGNVRFLLHRAKADVRKYKDYDIPSGKVLPIVSAVHLDN 416
 Db 321 GLEDVSMKFTRAVIYETSLATIVGVLRKTRDEINCYLLPKGRIVYVYREINYDA 380
 QY 417 SRYDOPLFPMPOOONNASSSGSFGSTGNNYMPFGGPRLCAGSELAKLEMAVFI 476
 Db 381 NLTEDDLIFPMRMKMSLESQ-----NSCFVRGGTRLCGKELGIVEISSFL 429
 QY 477 HHLVLFKNMELADDPFAPFVDFPNCGLPIRVS 510
 Db 430 HYFVTRYRMEIGDELMVFPVFAFKGFIHLRIS 463
 RESULT 14
 AAG11836
 ID AAG11836 standard; Protein; 461 AA.
 AC AAG11836;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 10708.
 XX
 KW Protein identification: signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EPI033405-A2.
 PD 06-SEP-2000.
 XX
 PE 25-FEB-2000; 2000SEP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134321.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0135124.
 PR 20-MAY-1999; 99US-0135491.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144086.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.

KM termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139753.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143342.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 19-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149375.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151530.
 PR 07-SEP-1999; 99US-0152363.

```
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0133758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 26.6%; Score 712; DB 21; Length 462;
Best Local Similarity 32.5%; Pred. No. 2,5e-54;
Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;

```
QY 12 LLLPSLSLFLILKRRN--KTRFNLPPGSGWPFLEGTIGYLRKYTTIGDFMQ 69
DB 2 MMILGLVLIIVLCYALLMNMQRYSKGLPGTGWMPFEGTTEFLKQGP----DFMK 56
QY 70 QHVSXGKYYRNSLFGPEPTIVASDAGLNRFILONEGRLEFECSPRSIGILGWSMLVLY 129
DB 57 NORLXGSEFFKSHIIGCEPTIVSMDBELNRYIILMNSKGLVAGYPOSMDILDTCNIAVH 116
QY 130 GDMHRDMRSISLNFSLHARLITILKDVERRHLFVLDWSQONSIFSAODEAKKFTFNMA 189
DB 117 GSHRLMRSGSLSLISPTMMKDHLLPKIDFMRNYLTCGWMDLEJVDIOEKTMAF-LSS 175
QY 190 KHMMSDPEEETEOUKKYYTFMKGVASAPLNLPGTAVHKALOSRAITLKFIERKMEER 249
DB 176 LQIAETLLKPEVEEYRTEFFKLVVGLTSPIDPTNYSRGVQARNNIDRLTTELMQER 235
QY 250 KLDKEEDOEVEVTEDEAEKMSKDHAKKORTDDDLGWLKHSN---LSTEQILDLI 305
DB 236 K-----ESGEFT-----DMGLYLMKKEDNRYLLTDKEIRQYV 268
QY 306 LSLFAGHETSSAIALAIFFLQACPKAVEELREHLEIARAKKEGESELNWDYKKMD 365
DB 269 VTIIVSGYETVETGMMALKYLDHPKALEELRREHLAIRKKRP--DEPLTLDIKSMK 326
```

```
QY 366 FTQCYINETLRIGNVYRFLHRAKADVRYKKYDIPSGKVLVPIVASHLDSRTDQPNLF 425
DB 327 FTRAVIFETSRLATIVNGVLRKRTTDLIELNGYLIPKGRIRIYYTREINVDLSLYEDPMIF 386
QY 426 NPMRMQOONNGASSGSGSFTWGNMYPFGGCPRLCAGSELAKEMAVFIHHLVLRKFM 485
DB 387 NPMRMKESLESK-----YFLIFGGVCLCQKELGISVSSFLHYFTYTKRM 435
QY 486 ELAEDQPPAFPEFVDFPNGLPYRS 510
DB 436 EENGEDKIMVFPFVSAPKGYHLKCS 460
```

Search completed: March 30, 2003, 12:04:18
Job time : 69 secs


```

DB 7 LLLLSIAGFL---LLLRTRYRMGLPPGSLGLPIGETFQILGAYKTENPEPFIDER 63
QY 72 VSKYKIRYNSLFGPEPTIVSADAGLNRFILQNEGRLEFCSYPSRIGIGLKMSMLVAGD 131
DB 64 VARYGSVFETHLFGPEPTIFSADPETRNFVLONEGKLEFCSYPASICNLGKSHLLMGCS 123
QY 132 MHRDRSISLNLSHARLRTILKDYERHTLVLDWMOONSIFSADDEAKKFTFNLMKX 191
DB 124 LKRRHSLTMSFANSIILKDHMLDIDRLVRENLDWSSRVLL--MEKAKITFELTVQ 181
QY 192 IMSDPEEETQOLKEEYTFMKGVASAPLNPCTAYHVALOSRATILKFIERKMEERL 251
DB 182 LMSFDPG-EMSSSLKREYLVIEGFSPLPLFSTTYKRAIAR-----RKVAEALT 232
QY 252 DIKEEDQEEVEEYKTEDEAMSKSDHVRKORTDDLLGWYLKSHNSTEQIIDLISLFA 311
DB 233 VVMMRRREEEGAE-----RKMDLAALLAADGFSDEIYDFVALLVA 278
QY 312 GHETSVAIALIFLQACPKAVEELREHLEIAAKKELGSELNMDYKMDTQCVI 371
DB 279 GETTSTMTTAVKFLTETPLALQLEHEKIRAKMSD--SYLSWSDYKSMPTQCV 336
QY 372 NETLRIGNVVRLHAKKALDYKGYDIPSGWKVLPVISAVHLDSRYDQPLFPMRMQ 431
DB 337 NETLVANIGVFRRAMTDEIKGKIPKMKVSSFRVAVHLDNHFADATFPMRMQ 396
QY 432 QONNGASSSGSFSWTGNN-YMPGGGPRCAGSELAKLENAVTHLVLKFNMLAED 490
DB 397 -----SNSVTTGNSNVTPEGGGPRCAGELARVALSVLHRLVTFGSVPAEQ 446
QY 491 DQFAFPVDFPNGLPRIYSR 511
DB 447 DKLVFPPTTRQKRYPIYKR 467

RESULT 2
US-08-622-166A-4
; Sequence 4, Application US/08622166A
; Patent No. 5952545
; GENERAL INFORMATION:
; APPLICANT: KONCZ, CSABA
; APPLICANT: MATHEUR, JAIDEEP
; APPLICANT: SZEKERES, MIKLOS
; APPLICANT: ALTMANN, THOMAS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
; TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS I
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622.166A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 0147-0153P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-622-166A-4

Query Match 36.5%; Score 978.5; DB 2; Length 472;
Best Local Similarity 41.9%; Pred. No. 5e-89;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLLSLILFLILKRRNRKTRPNLPKSGMPGIGETGYLKPYATATLGPDMOCH 71
DB 7 LLLLSIAGFL---LLLRTRYRMGLPPGSLGLPIGETFQILGAYKTENPEPFIDER 63
QY 72 VSKYKIRYNSLFGPEPTIVSADAGLNRFILQNEGRLEFCSYPSRIGIGLKMSMLVAGD 131
DB 64 VARYGSVFETHLFGPEPTIFSADPETRNFVLONEGKLEFCSYPASICNLGKSHLLMGCS 123
QY 132 MHRDRSISLNLSHARLRTILKDYERHTLVLDWMOONSIFSADDEAKKFTFNLMKX 191
DB 124 LKRRHSLTMSFANSIILKDHMLDIDRLVRENLDWSSRVLL--MEKAKITFELTVQ 181
QY 192 IMSDPEEETQOLKEEYTFMKGVASAPLNPCTAYHVALOSRATILKFIERKMEERL 251
DB 182 LMSFDPG-EMSSSLKREYLVIEGFSPLPLFSTTYKRAIAR-----RKVAEALT 232
QY 252 DIKEEDQEEVEEYKTEDEAMSKSDHVRKORTDDLLGWYLKSHNSTEQIIDLISLFA 311
DB 233 VVMMRRREEEGAE-----RKMDLAALLAADGFSDEIYDFVALLVA 278
QY 312 GHETSVAIALIFLQACPKAVEELREHLEIAAKKELGSELNMDYKMDTQCVI 371
DB 279 GETTSTMTTAVKFLTETPLALQLEHEKIRAKMSD--SYLSWSDYKSMPTQCV 336
QY 372 NETLRIGNVVRLHAKKALDYKGYDIPSGWKVLPVISAVHLDSRYDQPLFPMRMQ 431
DB 337 NETLVANIGVFRRAMTDEIKGKIPKMKVSSFRVAVHLDNHFADATFPMRMQ 396
QY 432 QONNGASSSGSFSWTGNN-YMPGGGPRCAGSELAKLENAVTHLVLKFNMLAED 490
DB 397 -----SNSVTTGNSNVTPEGGGPRCAGELARVALSVLHRLVTFGSVPAEQ 446
QY 491 DQFAFPVDFPNGLPRIYSR 511
DB 447 DKLVFPPTTRQKRYPIYKR 467

RESULT 3
US-08-724-466B-2
; Sequence 2, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; OPERATING SYSTEM: COMPAG, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724.466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667.546
; FILING DATE: June 21, 1996

```

ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466b-2

Query Match 18.8%; Score 503; DB 3; Length 492;
Best Local Similarity 27.6%; Pred. No. 1,5e-41;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

QY 7 HTLLPLLLPSLLSLFL-----ILKRRNRKTRFNPFGKSGWPLFGTIGYLPK 58
DB 4 YTLAVTFCTIVLPVLLFLAAVKLMEMLMIRVDPNCKSPPLPGTMGLPIGTIGTLQL-- 61
QY 59 YTATTLGDFMOQHYSKYKTRNSLFGEPITYVSADAGLNRIPLONEGRLEFCSYPRSIG 118
DB 62 ---LQRRKFLMKRQKCYIKTHLFGNPTVYMGADVNRQILLGEHLVSVOVPASVRT 118
QY 119 ILGKMSMLVVGDMHRD-----MRSISLNFSLHARLRTILKDYERHTLFVLDSW-QOONS 172
DB 119 ILGSDTSLNNGVGHKKKRAIMAFSDALEH-----YIPVIOQEVKSAIOEWLQDS 172
QY 173 IFSADAEKKTFFMLAKHNSMDPGEET--EQLKREYVFMKGVSAPIPLPGTAYHK 230
DB 173 CVLVYPEKKLMFRAMKILLEGPEQIKTDEQLVEAFEEKIKMLFSLPIDVPSGLYR 232
QY 231 ALQSRATILKIERKMERKLDIKEEDQEEVEVTEDEAEKSKDHYRKQRTDDLLGV 290
DB 233 GLRRAR---NFHSKIEENIRKRIODDNEQKYKALQLL---IENSRSDE----- 279
QY 291 LKHSNLSLEQILDLSLFGHETSSVAIALAIFFLQACPKAVEELREHELEIRAKKE 350
DB 280 ---PFSLQAKKEATLLEFGHETSTASTSLVFLGLNTEVQKVEE---VOEKVE 331
QY 351 LG---ESELNWDYKKMDFTQCVINETLRIGNVRFLLHRAKLDVRYKGYDIPSGWVL 406
DB 332 MGMTTPGKGLSMELLDQKYTCVIKETLRINPPVPGGFVALKTFELNGYQIPGMWVI 391
QY 407 PVISAVHLNDSRYDQPNLFNFWMOQONNGASSSGSFTWGNMYMPFGGPRICAGSE 466
DB 392 YSICDTHVADVFPNKEEFQPERFW-----SKGLEDSRP--NYIFGGGSRMCVGE 442
QY 467 LAKLEMAVFIHHLVLFKFMWELAEDDQ-----PFAFPVDFP 502
DB 443 FAKVLKIFLVELTQHCWILSNGPPTMKGTPTIYVDNLP 483

RESULT 4
US-08-882-164D-2
Sequence 2, Application US/08882164D
Patent No. 6306624
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Casseels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-2

Query Match 18.8%; Score 503; DB 4; Length 492;
Best Local Similarity 27.6%; Pred. No. 1,5e-41;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

QY 7 HTLLPLLLPSLLSLFL-----ILKRRNRKTRFNPFGKSGWPLFGTIGYLPK 58
DB 4 YTLAVTFCTIVLPVLLFLAAVKLMEMLMIRVDPNCKSPPLPGTMGLPIGTIGTLQL-- 61
QY 59 YTATTLGDFMOQHYSKYKTRNSLFGEPITYVSADAGLNRIPLONEGRLEFCSYPRSIG 118
DB 62 ---LQRRKFLMKRQKCYIKTHLFGNPTVYMGADVNRQILLGEHLVSVOVPASVRT 118
QY 231 ALQSRATILKIERKMERKLDIKEEDQEEVEVTEDEAEKSKDHYRKQRTDDLLGV 290
DB 233 GLRRAR---NFHSKIEENIRKRIODDNEQKYKALQLL---IENSRSDE----- 279
QY 291 LKHSNLSLEQILDLSLFGHETSSVAIALAIFFLQACPKAVEELREHELEIRAKKE 350
DB 280 ---PFSLQAKKEATLLEFGHETSTASTSLVFLGLNTEVQKVEE---VOEKVE 331
QY 351 LG---ESELNWDYKKMDFTQCVINETLRIGNVRFLLHRAKLDVRYKGYDIPSGWVL 406
DB 332 MGMTTPGKGLSMELLDQKYTCVIKETLRINPPVPGGFVALKTFELNGYQIPGMWVI 391
QY 407 PVISAVHLNDSRYDQPNLFNFWMOQONNGASSSGSFTWGNMYMPFGGPRICAGSE 466
DB 392 YSICDTHVADVFPNKEEFQPERFW-----SKGLEDSRP--NYIFGGGSRMCVGE 442
QY 467 LAKLEMAVFIHHLVLFKFMWELAEDDQ-----PFAFPVDFP 502
DB 443 FAKVLKIFLVELTQHCWILSNGPPTMKGTPTIYVDNLP 483

RESULT 5
US-08-882-164D-32
Sequence 32, Application US/08882164D
Patent No. 6306624
GENERAL INFORMATION:

```

APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckelt, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-32

```

```

Query Match 17.2%; Score 461; DB 4; Length 497;
Best Local Similarity 28.4%; Pred. No. 2.4e-37;
Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

QY 10 LPLLLPSL-----LSLLFLILLLK-----RRNRKTFNLPKSGSGMPFGETGYLK 57
DB 3 LPALLASALCTEVLPLLLFLAIAIKLMDLYCVSGRDSCLPLPPTGMPGPFGETL----- 58
QY 58 PYTATTIGDMOHOVSKYGIYRSNLFGEPTIVSADAGINRFLONEGRFLFCSPRSIG 117
DB 59 -QMWLQRRKFFLMKRRKTYFIKTHLFGRPYRVAGADVRRILLDGDLVSVHNPASVR 117
QY 118 GILGKWSMLVLDGMHDMRSISLNFSLHARLR--TILKVERHTLFLVDSMOO--NSIFS 175
DB 118 TILGAGCSLNLDSHKKQKVIYMAFSAREALQCYVLA-IAEVSCLQEWLSCGRRLL 176
QY 176 ADDEAKKFFNFNMAKIMSDP-----ETEQLKKEYVTMPKGVASAPLNIPGAYHKA 231
DB 177 VYPEKRLMFRIAMRLLGCEPQAGGDEQOLVEAFEMTRNLFSLPIDVPFSGLYRG 236
QY 232 LOSRAATILKFIKKME--RKLDIKEED--OEEEVKTEDEAEKSKSHVKKORTDDL 286
DB 233 VARNULIHARIEENIRAKTRRLQATEPDGCKDALQILIEHSWE-----RGERLDMO- 288
QY 287 LGMWLKHSNLTSTEOILDLILSLFAGHETSSVAIALAIFLOACPRAVEELREHLEIAR 346
DB 289 ---ALKQS--STE-----LLFGHETTASAATSLIYLDLHYHVLQKAVEEIKSKBL 335
QY 347 AKKEGESLFLANDYKMDFTQCVINETIRLCNVNRFLLRRAKLDKVRKYGYDIPSGWYVL 406
DB 336 LCKSNODKRLDDELTETLYKICVYIKETRLNLPVPVGGFRVALKTFELNGYQIPKGMVNI 395
QY 407 PYISAVHLDNSYVDGPNLFNFWRMQOONNGASSSGSSTSTGMNNTMPGGGPRCLCAGE 466

```

```

DB 396 YSICDTHVADIFTNKEEFPNDRFIVHPHEDASRFS-----FIPEGGLRSCVGR 446
QY 467 LAKLENAVEFIHMLVLFKNMELAE-----DDQPAFPFVDFP 502
DB 447 FAKILKIFTVELARHCQWQNLNGPPTMTSPYIVVDNLP 487

```

```

RESULT 6
US-08-724-466B-4
Sequence 4, Application US/08724466B
Patent No. 6063606
GENERAL INFORMATION:

```

```

APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckelt, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-4

```

```

Query Match 16.4%; Score 439; DB 3; Length 497;
Best Local Similarity 27.0%; Pred. No. 3.7e-35;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 10 LPLLLPSL-----LSLLFLILLLK-----RRNRKTFNLPKSGSGMPFGETGYLK 57
DB 3 LPALLASALCTEVLPLLLFLAIAIKLMDLYCVSGRDSCLPLPPTGMPGPFGETL----- 58
QY 58 PYTATTIGDMOHOVSKYGIYRSNLFGEPTIVSADAGINRFLONEGRFLFCSPRSIG 117
DB 59 -QMWLQRRKFFLMKRRKTYFIKTHLFGRPYRVAGADVRRILLDGDLVSVHNPASVR 117
QY 118 GILGKWSMLVLDGMHDMRSISLNFSLHARLR--TILKVERHTLFLVDSMOO--NS 172
DB 118 TILGAGCSLNLDSHKKQKVIYMAFSAREALQCYVLA-IAEVSCLQEWLSCGRRLL 176
QY 173 ISADDEAKKFFNFNMAKIMSDP-----GEEETDQKKEYVTMPKGVASAPLNIPGTA 227
DB 174 GLVYPEYKRLMFRIAMRLLGCEPQAGGDEQOLVEAFEMTRNLFSLPIDVPFSG 232
QY 228 YKALOSRAATILKFIKKMEERLIDIKEED-----QEEEVKTEDEAEKSKSDHVRKORT 282
DB 233 LYRGKARULIHARIEENIRAKTRRLQATEPDGCKDALQILIEHSWE-----RGERL 285
QY 283 DDLGLGWLKHSNLTSTEOILDLILSLFAGHETSSVAIALAIFLOACPRAVEELREHLEI 342

```

Db 286 DMQ-----ALKOS--STE-----LLEGGHETASATSLITYLGLYPHVLOKVEELK 331
 QY 343 E1ARAKKELGESSELMNDYKMKDFQCVINETLRLGNVRF1HRKALDKVARKGDIPSG 402
 Db 332 SKGLLCKSNQDNKLDMELEBQKTYIGCVIKETLRLNPPVGGFRVALKTEELNGYQIIPKG 391
 QY 403 WYVLPYISAVHLDNSRYDQPNLFNPRWQOONNGASSSGSGSFSTWGNVMPFGGGRPLC 462
 Db 392 WNVYISICDTHVAE1FTNKEEFNDRFSAPHPEDASRFS-----F1PFGGGLRSC 442
 QY 463 AGSE1AKLEMAVEF1HHLVKFNWELAE-----DDQPAFPFVDP 502
 Db 443 VKREFAKILK1FTVELARHCDWQMLNGPPTKTSPTVYVNDLP 487

RESULT 7
 US-08-882-164D-4
 ; Sequence 4, Application US/08882164D
 ; Patent No. 6306624

GENERAL INFORMATION:
 APPLICANT: Petkovich, P. Martin, White, Jay A.,
 APPLICANT: Beckett, Barbara R., Jones, Glenville
 TITLE OF INVENTION: Retinoid Metabolizing Protein
 NUMBER OF SEQUENCES: 43
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Blake, Cassels & Graydon
 STREET: Box 25, Commerce Court West
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5L 1A9
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
 COMPUTER: COMPAQ, IBM PC compatible
 OPERATING SYSTEM: MS-DOS 5.1
 SOFTWARE: WORD PERFECT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/882,164D
 FILING DATE: June 25, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/667,546
 FILING DATE: June 21, 1996
 APPLICATION NUMBER: 08/724,466
 FILING DATE: October 1, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Hunt, John C.
 REGISTRATION NUMBER: 36,424
 REFERENCE/DOCKET NUMBER: 50767/00010
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 863-4344
 TELEFAX: (416) 863-2653
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 497 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-882-164D-4

Query Match 16.4%; Score 439; DB 4; Length 497;
 Best Local Similarity 27.0%; Pred. No. 3.7e-35;
 Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 10 LPILLPSL-----LSLILPILK-----RRNKRTRNLPFGSGWFLGEGTGYLK 57
 Db 3 LPILLASLCTFVLPPLFLAATKLDLYCVSGDRSCALPFPGTMGPFEGETL----- 58
 QY 58 PYATATLGFDMQOHSKRYKYSNLFGEPTIYSADAGLNRF1LQNEGR1FEGSGYRSTG 117
 Db 59 -QWVLGRKFLQMKRRKRYGTYKTHLFGRTYAVMGADNVRILLDDRLVSYHWPASVR 117
 QY 118 GILGKMSLVYVGDMDRMSISLNFLSHARL---TILKDVERTLFLVDSWQO--NS 172

Db 118 TILSGGCSNLDHSHKORRVIMAFSREALCEYVPVITEVSS-----LEQWLSGGER 173
 QY 173 IFSADDEAKKFTFNMAMKIMSMP-----GEEETEOUKREYVFMGVASAP1NLGTA 227
 Db 174 GLVYPEYKRLMFR1AMN1LIGCEPQLAGDDSE--QOLVAFEEKTRNLSLPIDVPSG 232
 QY 228 YHK1QSAT1LTKFERMEERKLDIKEED-----QEEEVKTEDEAMSKSDHVRKQRT 282
 Db 233 LYRGKAKRNL1HAR1EQIRAK1G1GRASEGQCKD1QL1L1HBSW-----RERL 285
 QY 283 DDDLGLWV1KHSN1TEQ1IDL1SL1FAGHETSVA1ALA1F1LQ1CPRAVE1REBH 342
 Db 286 DMQ-----ALKOS--STE-----LLEGGHETASATSLITYLGLYPHVLOKVEELK 331
 QY 343 E1ARAKKELGESSELMNDYKMKDFQCVINETLRLGNVRF1HRKALDKVARKGDIPSG 402
 Db 332 SKGLLCKSNQDNKLDMELEBQKTYIGCVIKETLRLNPPVGGFRVALKTEELNGYQIIPKG 391
 QY 403 WYVLPYISAVHLDNSRYDQPNLFNPRWQOONNGASSSGSGSFSTWGNVMPFGGGRPLC 462
 Db 392 WNVYISICDTHVAE1FTNKEEFNDRFSAPHPEDASRFS-----F1PFGGGLRSC 442
 QY 463 AGSE1AKLEMAVEF1HHLVKFNWELAE-----DDQPAFPFVDP 502
 Db 443 VKREFAKILK1FTVELARHCDWQMLNGPPTKTSPTVYVNDLP 487

RESULT 8
 US-09-615-192A-405
 ; Sequence 405, Application US/09615192A
 ; Patent No. 6410718

GENERAL INFORMATION:
 APPLICANT: Bloksberg, Leonard N.
 APPLICANT: Havukkala, Ilkka
 TITLE OF INVENTION: Materials and Methods for the
 FILE REFERENCE: 11000.1003CAU
 CURRENT APPLICATION NUMBER: US/09/615,192A
 FILING DATE: 2000-07-12
 PRIOR APPLICATION NUMBER: US 08/975,316
 PRIOR FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: US 08/713,000
 PRIOR FILING DATE: 1996-09-11
 PRIOR APPLICATION NUMBER: US 09/169,789
 PRIOR FILING DATE: 1998-10-09
 NUMBER OF SEQ ID NOS: 405
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 405
 LENGTH: 529
 TYPE: PRT
 ORGANISM: *Escherichia coli*
 US-09-615-192A-405

Query Match 11.3%; Score 302.5; DB 4; Length 529;
 Best Local Similarity 21.9%; Pred. No. 1.7e-21;
 Matches 117; Conservative 104; Mismatches 204; Indels 109; Gaps 18;

QY 11 PILLPSLILSLILKRRNKRTRNLPFGSGWFLGEGTGYKPYATATLGDPMQO 70
 Db 22 PMLLSVLPPLFLGLVLR1LRKRPF--PPGRG1PV1GNML-----MMSELTRH 70
 QY 71 HVS-----KYGK1YRSNLFGEPTIYSADAGLNRF1LQNEGR1FEGSGYRSTG 116
 Db 71 GLASLAKRYG1FHLRMGL1HVA1VAVSP1VARQV1QVH1DG1FSN1RPA1TA1SL1Y1D1RAD 130
 QY 117 GGLGKMSLVYVGDMDRMSIS--LNF1SHARL1FT1LK1DVERTL1FLVDSWQO1N1S1FS 175
 Db 131 -----MAFAHYG1PFRW1MR1L1CYMK1LFS1RRK-----AS1WE-----S 162
 QY 176 ADEAK-----KFTFNM1AMK1IMS1MPGEEETEOUKREYVFMGVASAP1NLGTA 214
 Db 163 VADVD1M1V1R1V1AG1SG1R1V1N1G1ELV1ELTRD1IT1RA1FG1S1TEG--ODE1F1S1LQ1ERSK 221

QY 215 --GVVSAPINLEGTAV-----HKALOSRATILFKIERKME---ERKIDIEEDOEDEE 262
DB 222 LKGAENIADPIFYLYLWIDPOGLTARLYVKAROSIDGFIHIDHMDKKNKSSGGGDOD 281
QY 263 VKTE--DEAEMSKSDHVRKQRTDDDLGVLKHS--NLSTEOILDLILSTLFGHETSVA 319
DB 282 VTDWVDLLAAYSDPAKNESD-----LQNSIRLIDNKAITMDVEGTEVASA 335
QY 320 IALAIFFLOACPRAVEELREHLEIARAKKEGSELMNDYKMDPQCVINETLRLGN 379
DB 336 IEMAAELKRSPEDLKQVQELADVGLDRVYES-----DEKILTYLKCCLETRLRHP 390
QY 380 VVRFLHRKALKVRKKGDIPIGSKVLYPIYSAVHLDNSYDDPNIENPWRQOONNGASS 439
DB 391 PIPLLHETAEADAVISGYRIPARSRVIMAAIGRDPGSGTEPDKFKPSRFL----- 443
QY 440 SGGSGFSTGNVMPFGGPRLCAGSELAKLEMAVFIHILVYKFMLEADDP 493
DB 444 SGMPDYKSGNEFIPIFGSGRSCPGKQGLYALDMAVAHLLHCTWELPDGMP 497

RESULT 9

US-08-991-677-4
; Sequence 4, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent L
; APPLICANT: Carraway, Daniel T
; APPLICANT: Smeltzer, Richard H
; TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
; FILE REFERENCE: 50617
; CURRENT APPLICATION NUMBER: US/08/991,677A
; CURRENT FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: US 60/033,381
; EARLIER FILING DATE: 1996-12-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Liquidambar styraciflua
US-08-991-677-4

Query Match 10.8%; Score 289; DB 4; Length 511;
Best Local Similarity 22.0%; Pred. No. 3.6e-20;
Matches 121; Conservative 95; Mismatches 196; Indels 138; Gaps 21;

QY 6 HTLLPL-LLPLSLLLFLLLLRNRKTRFNLPGKSGMPFLGRTIGYLAQPTATTL 64
DB 6 HEADLPPLPTLFTPLPLLLGLVSR--LRQLPYRPGKGLPVIGNML-MMDQLTHRGL 62
QY 65 GDFMOOHVSKYKIRSNLFGERTIVSADAGINRFLONEGRLEF-----CSYPR 114
DB 63 AKLANQ---YGLFLHKKMGFLHMAVSTPPDARQVLOQDNIFSNRPATIASLVTYR 118
QY 115 STIGILGKWSMLVYGDHNRDRSIS-LNFIASHARLRTLLKDVERTLFLVDSMOQNSI 173
DB 119 AD-----MAFAHYGPRWQRKLCVMKLFSSRR-----AESWE----- 151
QY 174 FSAODE-----AKKFTFNLMAKHIMSMDGE-----ETE 203
DB 152 -SVRQVDASAVRNASNIGSTIVNIGELVAFALTKNITTYRAAGTISHEDODEVAILOES 210
QY 204 QL-----KREYVTFMKGV---VSADPLNPGTAYHKALOSRAITLFIKIERKMERKLDIKE 255
DB 211 QLFQAFNLADFTPLMKWYPOGINVRLN-----KARGLDQFIKSIIDH---IQK 257
QY 236 EDQEEBEVKT-----DEAEMSKSDHVRKQRTDDDLGVLKHSNLSSTEOILDLI 305
DB 238 GSKNESEVDVMDLLAFYGEAEKAVSESDD-----LQNSIKILTKNIKAI 303
QY 306 LSLFLAGHETSSVALAIFLQACPRAVEELREHLEIARAKKEGSELMNDYKMD 365

DB 304 MDVFMGTEETVASALEMATELTKSPEDLKQVQELAAVVGIDRVEEK-----DEKILT 358
QY 366 FQOCVINETLRGNVVRFLHRKALKDVRKKGDIPIGSKVLYPIYSAVHLDNSRYDDPNI 425
DB 359 YLKYVLEKVLRLHPRPLILLHETAEADVAEYGYIIPAKSKVMINACAGDKKSNADPDE 418
QY 426 NPMRQOONNGASSSGSGFSTWGN--YMPFGGPRLCAGSELAKLEMAVFIHILVYK 483
DB 419 RRSRFLK-----GVPDFK--GNNEFIPIFGSGRSCPGKQGLYALETVAHLLHCF 469
QY 484 NMELEADDP 493
DB 470 TWELPDGMP 479

RESULT 10

US-08-948-564-6
; Sequence 6, Application US/08948564
; Patent No. 6121512
; GENERAL INFORMATION:
; APPLICANT: Siminszky, Balazs
; APPLICANT: Dewey, Ralph E.
; APPLICANT: Corbin, Frederick T.
; TITLE OF INVENTION: No. 6121512a1 Cytochrome P-450 Constructs and
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
; NUMBER OF SEQUENCES: 23
; CORRESPONDING ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 6121512th Carolina
; COUNTRY: USA
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,564
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 5051-409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-948-564-6

Query Match 10.4%; Score 279.5; DB 3; Length 513;
Best Local Similarity 22.9%; Pred. No. 3.2e-19;
Matches 118; Conservative 82; Mismatches 216; Indels 99; Gaps 17;

QY 15 LPSLLSLFLLLLRNRKTRFNLPGKSGMPFLGRTIGYV---KPYATATLGDPMOON 71
DB 15 LAFPLISGLIF--FLKOKSKKFNLPDGPWPRIYGNLFQVARSKPPF-----EYVNV 67
QY 72 VSKYKIRSNLFGERTIVSADAGINRFLONEGRLEFECSTYR----- 114
DB 68 RLKYSIFLTKGTRMTIITDAKLVHEAMIOKQATYATRPENPTRTIFSENKFTVNA 127
QY 115 STIGILGKWSMLVYGDHNRDRSISLNPLSHARLRTLLKDVERTLFLV-----DSW 168
DB 128 TYGPV---WKSLL-----RRNVQNMLSSTRUKE--FRSVRDNAQDKLINRLKDEAE 173

[illegible]

```

RESULT 11
US-09-144-367-2
: Sequence 2, Application US/09144367
: Patent No. 6432639
: GENERAL INFORMATION:
: APPLICANT: Lichter, Jay
: APPLICANT: Guido, Marco
: TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
: FILE REFERENCE: SEQ-12P
: CURRENT APPLICATION NUMBER: US/09/144,367
: CURRENT FILING DATE: 1998-08-31
: PRIOR APPLICATION NUMBER: 60/058,612
: PRIOR FILING DATE: 1997-09-10
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 503
: TYPE: PRT
: ORGANISM: H. sapiens
US-09-144-367-2

```

```

Query Match 273; 10.2%; Score 273; DB 4; Length 503;
Best Local Similarity 12.4%; Pred. No. 1,4e-18;
Matches 120; Conservative 101; Mismatches 222; Indels 92; Gaps 21

OY 9 LRLPLLLSLSLLFLILIKRRNRKRENL---PGKSGMFLGETIGLYKPYATTL 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 LRLPLAMETWLLLAALSVLLLYGTHSHGLFKKLGIPGTPPLPGLNIIISYHGFQCM-- 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 65 GDFMOAHVSKYKIRSNLFGPTIYSAAGLNRFLLQNEGRFLFEGSYPRSTIGILG--K 122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 --EDMECHKKKYKRWGFIYDGOQPVLAIDPDMIKIYLVKE--CYSYFTNRPRPGPFGPK 115
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 123 WSMVLVYGDMDHDMKS--ISLNLFLSH-----ARLRTILIKDVERH-----TLFVYL 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 SAISIAEDDEEMRLSLISLPFTSGKLEMPVLIAGYGVVLVRLNRRRAETGKPVTL-- 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 166 DSMQONSFEA--QDEAKKFTFNLMAAKHIMS--MDPGEETEDQAK--EYV--TFPKGVYSP 220
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 ----KIDVFGAISMVUTISTSGVINDISLNNODPPEVETKKILRDLDPFLSLTVP 227
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 221 INLPSTAYHKAQSRAITILKTERKMEERKDLIKEEDQEEVEKYTDEAEAMSKSDHVRQ 280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 FLIPLEVLNICVPPREVYNEL--RSYKMKMSRLDEDQKNHY---DFQL--MDSQNSK 282
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 281 RYDDDLGLGVLKHSNLSQTQIDLLISLLFAGHETSSVAIALAIFLQACRAVEELREE 340
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      283 ETE-----SHKALDELVAQSIIFIFAGYEFTSSVJSFIMELATHPHDVQKQDE 334
QY      341 HLEIARAKKEIGESELNDNDYKQKDFQCVINETFLRLGNVRETLHKKALDVRKGYDIP 400
Db      335 -IDAVLPNK---APPTVDYLOMEYDVMVNETLRLFLPIAMREBRCKKQVEINGMFIP 369
QY      401 SGWVYLPVISAVHLDNSRYDDPNLEFNPRMOOONNGASSSSGGSFTWGN-----YMPF 455
Db      390 KGWVVMIPISVYLHBDPKWTPEKFLPELFERFEKKK-----DNIDPYIYTPF 435
QY      456 GGGPRLCGSELAKLEMAVFTHHLYLVLENWELADDDOPAFVPYDFPNGALPIYRS 510
Db      436 GSGPRNCIGMFPALMNKIALIRVLQNFSEFCKETO-----IDLKIS 478

```

```

, RESULT 12
, US-09-126-420A-18
, Sequence 18, Application US/09126420A
, Patent No. 6376753
, GENERAL INFORMATION:
, APPLICANT: BATARD, YANNICK
, APPLICANT: ROBINEAU, TIBURCE
, APPLICANT: DUREST, FRANCIS
, APPLICANT: WERCK-REICHART, DANIELE
, APPLICANT: DIDIERJEAN, LUC
, TITLE OF INVENTION: PORIFIED CYTOCHROME P450 CYP76B1 FROM HELIANTHUS
, TITLE OF INVENTION: TUBEROSUS AND ITS APPLICATIONS AS BIOCATALYST IN
, TITLE OF INVENTION: PARTICULAR FOR THE DEGRADATION OF ENVIRONMENTAL
, TITLE OF INVENTION: POLLUTANTS AND FOR ALTERING THE RESISTANCE OF PLANTS S
, FILE REFERENCE: 03715_0032
, CURRENT APPLICATION NUMBER: US/09/126,420A
, CURRENT FILING DATE: 1998-07-30
, PRIOR APPLICATION NUMBER: 60/054,351
, PRIOR FILING DATE: 1997-07-31
, NUMBER OF SEQ ID NOS: 27
, SOFTWARE: PatentIn Ver. 2.1
, SEQ ID NO 18
, LENGTH: 504
, TYPE: prt
, ORGANISM: Solanum melongena
, US-09-126-420A-18

```

[illegible]

QY 405 VLPLVSAVHLNLSRQDQNLNPPNRMQOONNGASSSSGSPSTGNVY --MPEGCGRLC 462
 Db 398 VLVNMAIGRDPEDCDPMSPKPRFL-----GSKIDVQGYHGLPFGAGRMC 447
 QY 463 AGSELAKLEMAVFTHLVLYKRNWELAEDDQ 493
 Db 448 VGLPLGRHMMHFAUGSLDLREFEWELPDGVSP 478

RESULT 13
US-09-499-302A-2
; Sequence 2, Application US/09499302A

```

1  GENERAL INFORMATION:
2  APPLICANT: BOUNG-JUN, OH
3  APPLICANT: MOON, KYUNG KO
4  APPLICANT: YOUNG, SOON KIM
5  TITLE OF INVENTION: A CYTOCHROME P450 GENE HIGHLY EXPRESSED IN THE
6  TITLE OF INVENTION: INCOMPATIBLE INTERACTION
7  FILE REFERENCE: 10324/P6443050
8  CURRENT APPLICATION NUMBER: US/09/499,302A
9  CURRENT FILING DATE: 2000-02-07
10 NUMBER OF SEQ ID NOS: 10
11 SOFTWARE: PatentIn Ver. 2.1
12 SEQ ID NO 2
13 LENGTH: 502
14 TYPE: PR1
15 ORGANISM: Capsicum annuum
16 US-09-499-302A-2

```

Query Match	10.0%;	Score 268.5;	DB 4;	Length 502;
Best Local Similarity	21.5%;	Pred. No. 3.9e-18;		
Matches 114;	Conservative 100;	Mismatches 206;	Indels 109;	Gaps 18;

```

QY 17 SLUSLLEF---ILKRNKRNTRPNPPGSGMPELEETIGLYKPYATLIGF---MQ 69
Db 7 NLVAFLEFLSSIIILLKKWKQKQKQLPRPGMKPLFISL-----FLVAVGRLPHNGLK 60
QY 70 QHVSXKGYKTYNSNFGEPTIYSADAGINRFLONEGRLEFECSTPRISGILL----- 120
Db 61 QLARLYGRLMLRLGELPTIYIISPPRAKVELKHNDAFATRFRLVYADIVHYDSDIDAF 120
QY 121 ---GKMSLVLYGDHNRDMRSISNPFASHARLTITLKDVBRHFLFYDS---WQNSIF 174
Db 121 SPYEGMT-----KQIRKICLLELLSKM-KFESSIRODELSMMSISRTMRFV 170
QY 175 SAODEAKFTPNLMAKHIMSDPCEEBETBOLKEITYTPMKGVVASP----- 220
Db 171 NLTKRIFMFTSSVYCSRSLGICIDDO-----KLIFMEBIIISLTGFSIADFFPTWM 224
QY 221 LNLPGTAVHKLQSRATITKIERKMEERKIDKEEJOEEYEVTEDEAE---MSKSDH 276
Db 225 LHVDGSGSTRLLKHKRKIDELLEHVAVHEKONRADGCKGGEFEGEDLIDYLKVRSGE 284
QY 277 VRKORTDDLLGWLKHSNLSSTEOIDLILSLFAGHETSSVAIALAIFLQACPAVEE 336
Db 285 VOISITDND-----IKSLIDVMSFAGSETISSTIIMALAMMKKPSYIAK 329
QY 337 LRSEHLEIARAKKELGSELSMNDQYKMDPTQOCINTELSLGNVRL-HRKALKDVRXK 395
Db 330 AQAEEVQYLKERR--GROQIDIDELK---YKULVKTETLMNHPITLILVRECKMKDKID 384
QY 396 GYDIPSGMKVLPVISAHLNRSRDQPNLEFNPMWQOONGASSGSGSFSTGNNMYP 455
Db 385 GYNIPKPRFVIVNMAIGRDPESMDDESPSPERF--ENSGSVDFLGS-----HQDITP 436
QY 456 GGGFRLCAGSELAKLEBAVFIHNLVYLFENNELAEDDQPAFAPFVDPNCPAR 506
Db 437 GAGRRICG-----MKEG--LANVGQPLQOLLHYEDRKLP 469

```

RESULT 14

US-09-126-420A-17
; Sequence 17, Application US/09126420A
: Patent No. 6376753

```

1 GENERAL INFORMATION:
2 APPLICANT: BATARD, YANNICK
3 APPLICANT: ROBINEAU, TIBURCE
4 APPLICANT: DORST, FRANCIS
5 APPLICANT: WERCK-REICHART, DANIELE
6 APPLICANT: DIDIERJEAN, LUC
7 TITLE OF INVENTION: PURIFIED CYTOCHROME P450 CYP76B1 FROM HELIANTHUS
8 TITLE OF INVENTION: TABEROSUS AND ITS APPLICATIONS AS BIOCATALYST IN
9 TITLE OF INVENTION: PARTICULAR FOR THE DEGRADATION OF ENVIRONMENTAL
10 TITLE OF INVENTION: POLLUTANTS AND FOR ALTERING THE RESISTANCE OF PLANTS S
11 FILE REFERENCE: 03715.0032
12 CURRENT APPLICATION NUMBER: US/09/126,420A
13 CURRENT FILING DATE: 1998-07-30
14 PRIOR APPLICATION NUMBER: 60/054,351
15 PRIOR FILING DATE: 1997-07-31
16 NUMBER OF SEQ. ID NOS: 27
17 SOFTWARE: PatentIn Ver. 2.1
18 SEQ. ID NO 17
19 LENGTH: 467
20 TYPE: PRN
21 ORGANISM: Solanum melongena
22 US-09-126-420A-17

```

Query Match	10.0%	Score 268	DB 4	Length 467
Best Local Similarity	23.2%	Pred. No.	3.9e-18	
Matches 113	Conservative 80	Mismatches 182	Indels 112	Gaps 18

QY	46	WPFLEGTIGYUKPYATATLLDGFMOOH-VSKGKTYRNLNCEPTIVADAGLNFILONE	104
Dd	28	WLKJGGSMNTMYIOTJANSASLELFRNHVS-----ESDRIVDVNLAHNY-----	72
QY	105	GRLEECSTPNSIGTILGKWS-LVVGDMHDMRST-SLNFSLHRL-RYLLKDERHT	161
Dd	73	-----KGSMLAPYGNMYFNSRRICTVEMFVHKRINETNTINIOESVDK	115
QY	162	LEVLDSMOONSIFSAOD-EAKRF-----TFNLAKHJMSDPCGEETEOJKEVTEMG	219
Dd	116	MLRIDEKKAASSGGGGEGIEYTRMFLASFPMVGNMFLSKDLYTDPPSKGSEFFNMIG	175
QY	216	VVSAPLNLPGTA-----YKALQ-----SRATLKEFERKMEBRKIDIKEED	257
Dd	176	IMEW-AGVPNLSIDIFPCLKMDVOGLRKKRERDMGKRETKKEIERIERBK-----	227
QY	258	QEEBEVTEDEAKMSKSDVHKORTDDLLGMY-----KRSNLTSTOJDLJLSL	309
Dd	228	-----KGE-----KNSINDLLDVLIDFEGSGKDEPDKLSEDEIYIILEMF	263
QY	310	FAGHETSSVALAIIFPLQACPKAVELEREHLEIARAKELGESELNMDYKMDTQC	369
Dd	270	LACETITSSSEVMTLELLRHPOAKMAKYLEIILVIGPNKKFECCID-----SLPYQA	324
QY	370	VINETLRLGNVRL-HRKALVDVRYGYDIPSGWKYLPVISAHJLDSRYPDQNLCPNW	428
Dd	325	VLRQQLRHPPLLRKALQDTRKEGYDIPKTOYLVANMAIGRDPREYWDNFEKPE	384
QY	429	RMQOONGGASSSGSSEFSTWGNNT--MPFGGPRILCAGSELAKEAMAVTIIHJVLKENNE	486
Dd	385	RFLR-----SKVDVGQNYELIPFGAGRMCGVGLRGRMHMFTFGSLLHERDWE	433
QY	487	LAEDOP 493	
Dd	435	LPHNVSP 441	

RESULT 15
US-09-126-420A-26
; Sequence 26, Application US/09126420A
; Patent NO. 6376753
; GENERAL INFORMATION:
; APPLICANT: BATARD, YANNICK

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2003, 12:08:26 ; Search time 51 Seconds

(without alignments)
590.691 Million cell updates/sec

Title: US-09-502-426a-2

Perfect score: 2681

Sequence: 1 MPEDEHHTLLPLLLPSLLS.....FAFPVDFPNCPIRVSRIIL 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: Published Applications-AA:*

1: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppa/PC7_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppa/PC7US_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	292	10.9	516	9 US-10-067-534-3	Sequence 3, Appli
2	290.5	10.8	514	10 US-09-947-027-4	Sequence 4, Appli
3	290.5	10.8	514	12 US-10-091-009-4	Sequence 4, Appli
4	289	10.8	511	10 US-09-796-256a-4	Sequence 4, Appli
5	285.5	10.6	502	10 US-09-739-254-6	Sequence 69, Appl
6	285.5	10.6	502	10 US-09-904-615-6	Sequence 69, Appl
7	275	10.3	496	9 US-10-067-668-10	Sequence 10, Appl
8	275	10.3	496	10 US-09-945-301-4	Sequence 4, Appli
9	272	10.1	470	9 US-10-067-668-12	Sequence 12, Appl
10	256.5	9.6	520	10 US-09-992-901-2	Sequence 2, Appli
11	251	9.4	515	10 US-09-796-138-19	Sequence 19, Appl
12	251	9.4	515	10 US-09-909-903-19	Sequence 19, Appl
13	251	9.4	544	9 US-10-067-668-8	Sequence 8, Appli
14	249	9.3	503	10 US-09-796-138-18	Sequence 18, Appl
15	249	9.3	503	10 US-09-909-903-18	Sequence 18, Appl
16	249	9.3	503	10 US-09-957-997-3	Sequence 3, Appli
17	249	9.3	509	10 US-09-817-184-4	Sequence 4, Appli
18	247.5	9.2	493	9 US-10-103-520-2	Sequence 2, Appli
19	247.5	9.2	505	10 US-09-910-689-307	Sequence 307, App

20	247.5	9.2	505	12 US-10-010-742-307	Sequence 307, App
21	246	9.2	509	9 US-10-028-072-108	Sequence 108, App
22	246	9.2	509	9 US-10-121-049-108	Sequence 108, App
23	246	9.2	509	9 US-10-123-904-108	Sequence 108, App
24	246	9.2	509	9 US-10-140-470-108	Sequence 108, App
25	246	9.2	509	9 US-10-175-746-108	Sequence 108, App
26	246	9.2	509	9 US-10-176-918-108	Sequence 108, App
27	246	9.2	509	9 US-10-176-921-108	Sequence 108, App
28	246	9.2	509	9 US-10-137-865-108	Sequence 108, App
29	246	9.2	509	9 US-10-140-474-108	Sequence 108, App
30	246	9.2	509	9 US-10-142-431-108	Sequence 108, App
31	246	9.2	509	9 US-10-143-114-108	Sequence 108, App
32	246	9.2	509	9 US-10-140-002-108	Sequence 108, App
33	246	9.2	509	9 US-10-142-419-108	Sequence 108, App
34	246	9.2	509	9 US-10-123-262-108	Sequence 108, App
35	246	9.2	509	9 US-10-142-423-108	Sequence 108, App
36	246	9.2	509	9 US-10-121-050-108	Sequence 108, App
37	246	9.2	509	9 US-10-141-755-108	Sequence 108, App
38	245.5	9.2	508	9 US-09-925-299-905	Sequence 905, App
39	245.5	9.2	508	10 US-09-925-299-905	Sequence 905, App
40	244.5	9.1	524	9 US-09-992-598-264	Sequence 264, App
41	244.5	9.1	524	9 US-09-989-293A-264	Sequence 264, App
42	244.5	9.1	524	9 US-10-063-547-54	Sequence 54, Appl
43	244.5	9.1	524	9 US-09-989-735-264	Sequence 264, App
44	244.5	9.1	524	9 US-09-980-444-264	Sequence 264, App
45	244.5	9.1	524	9 US-09-989-730-264	Sequence 264, App

ALIGNMENTS

```
RESULT 1
US-10-067-534-3
: Sequence 3, Application US/10067534
: Publication No. US20020187538A1
: GENERAL INFORMATION:
: APPLICANT: Essenberg, Margaret K.
: APPLICANT: Chen, Xiao-Ya
: APPLICANT: Luo, Ping
: TITLE OF INVENTION: cDNA Clone of (+)-Delta-Cadinene-8-Hydroxylase Gene from Colt
: FILE REFERENCE: 006602-113
: CURRENT FILING DATE: 2002-02-07
: PRIOR APPLICATION NUMBER: US 60/267,160
: PRIOR FILING DATE: 2001-02-07
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 516
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
US-10-067-534-3

Query Match      10.9%; Score 292; DB 9; Length 516;
Best Local Similarity 23.6%; Pred. No. 7.8e-16;
Matches 120; Conservative 92; Mismatches 220; Indels 76; Gaps 17;

12 LLLLSLLFLILKRRNRKTRFNLPPKSGMFFLGEGTIGYLPYATTLGDFMOOH 71
20 IYLLTVSEIILWTF-----KSPQSLPPGPGRLPIVG-NLPFLDPDLHTFANLAOSH 73
72 VSKRYGKYSNLFGEFTVSADAGLNRFLLQNGRLF-----CSYPSIGIGLCKWSM 125
74 ----GIFKLTNGSKLTIYVNSFSLAREILKQDIDFNSRDVPLNGRAATYGGIDIVWPP 129
126 LVLVGMDHMDMSIS-LNLSHARLFT-----ILKQVREHTLFVLDWMOONSIFSADDEK 181
130 ---YGAEWQOLKIKICVLKLSRRTLDSEYELRKRERTRLYLEGGKQSPVYKVDQLF 186
182 KFTFNIMARKHNSMDGEEETDLKREYTFMKGVVS-----APLNL 223
187 LTMNMLTMMNMGSGYKADKEMESVGTFF-----KGVLSITRLISEPHVSDFPMLARFDL 242
```

[illegible]

```

RESULT 2
US-09-947-027-4
: Sequence 4, Application US/09947027
: Patent No. US20020124281A1
: GENERAL INFORMATION:
: APPLICANT: Chiang, Vincent Lee C.
: APPLICANT: Li, Laiyeng
: TITLE OF INVENTION: METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND COMPOSITIONAL CONTENT OF CELLULOSE IN PLANTS
: FILE REFERENCE: 066040-9718
: CURRENT APPLICATION NUMBER: US/09/947,027
: CURRENT FILING DATE: 2001-09-05
: PRIOR APPLICATION NUMBER: 60/230,086
: PRIOR FILING DATE: 2000-09-05
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 514
: TYPE: prt
: ORGANISM: aspen populus tremuloides
US-09-947-027-4

```

```

Query Match      10.8%; Score 290.5; DB 10; Length 514;
Best Local Similarity 22.6%; Pred. No.1e-15;
Matches 123; Conservative 98; Mismatches 230; Indels 93; Gaps 19;

QY      11 PLLPLSLSLLEPLILKLRNRKRTRENLPPQSGSWPFLGETIGYGLKRYATTTGLDFFMQO 70
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      12 PMSLFYLIYSSLFFEGGLSLRLR--RLPYPPGPKPLPLVG--SMHMDITHRGLAKLAKQ 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      71 HVSKGIKIRSNLGEPTIVSADAGLNRFIIQNEGRLE-----CSYPSIGTIL 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      69 ----YGLFHRMGVLIHWVTVSSPEIARQVLQYQDNITSNRPANIAISLYLRDAD----- 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121 GKMSMLVIVGDMHRMRSIS--LNLISHARLT-----ILKDVRRH-----T 161
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      121 ---MAFAHYGFPMQRKLCYVKLFESRKRAESWESVRDVSMLKTVANIGKPYNLGEL 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      162 LFLVLDSDQONSTFFSQDQAKFTFNLMAKHMSMDPGEEFLQOLKKEYVTMKGCVASAPL 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      178 IFTL---TMNITTYRAAFGAK-----NEGQDEFIKIQEF--SKLGFANMSD 219
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      222 NLPGATVAKALQSRATIIKFTIERKMEERKLD--IKEEQEEVEKTEDEAEKMSD----- 275
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      220 FIPWGLGWIDPQGLSARIVK--ARKALDRFIDSIIIDHHIQKRNKFSDEADFDWDMLA 277
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      276 ----HYKQRFDDDLGLGWLKHSNLSLSTEQIIDLLISLFLAGHETSSVAIALAIFFLQACP 331
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      278 FYGEEARARVSDSDL---QKAISLTKNIKAIIMDVAFGTEVAASAIEWMAELMKSP 333
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      332 KAVELREHLEIARAKKELGESLNMDDYKKMDFTQVINTETLRIGNVAFRLHKKALKD 391

```

Db	334	EDOKVQCELEAVGLERVEESID-----KLTFLCALKEFLRMPPPLLLHETSED	3688
Oy	392	VRYKGYDIPSGMKVLPYSAVHLNLSRYDNPENPKRMQOQNNGASSGSGSPSTGNN	451
Db	369	AEVAGYFLPKQTRVINAAYAGRKNSKNEDDAKPFRFLK-----PCVPDK--GNH	4399
Oy	452	--YMPFGGPRLCAGSELAKLEMAVFTHHLVLYKENWELAEADDPFAPFVDFPGLPIRY	509
Db	440	FEFLFEGSGRRSGCMQGLYTLDAVAHLHCETWELPGMKPSELDMDMGFLAPRA	4699
Oy	510	SRIL 513	
Db	500	TRLV 503	

```

1      RESULT 3
2      US-10-091-009-4
3      Sequence 4, Application US/10091009
4      Patent No. US20020138870A1
5      GENERAL INFORMATION:
6      APPLICANT: Chiang, Vincent Lee C.
7      APPLICANT: Li, Laiyeng
8      TITLE OF INVENTION: METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND
9      TITLE OF INVENTION: COMPOSITION, AND
10     TITLE OF INVENTION: CELLULOSE CONTENT IN PLANTS
11     FILE REFERENCE: 066040-9718
12     CURRENT APPLICATION NUMBER: US/10/091,009
13     CURRENT FILING DATE: 2002-03-06
14     PRIOR APPLICATION NUMBER: 09/947,027
15     PRIOR FILING DATE: 2001-09-05
16     PRIOR APPLICATION NUMBER: 60/230,086
17     PRIOR FILING DATE: 2000-09-05
18     NUMBER OF SEQ ID NOS: 14
19     SOFTWARE: PatentIn version 3.0
20     SEQ ID NO 4
21     LENGTH: 514
22     TYPE: PRF
23     ORGANISM: aspen populus tremuloides
24     US-10-091-009-4

```

Query Match	10.8%	Score 290.5	DB 12	Length 514
Best Local Similarity	22.6%	Pred. No. 1e-15		
Matches 123	Conservative	98	Mismatches 230	Indels 93
			Gaps	19
QY	11	PLLLPILSLILSLLEITILKRRNRKTRFNLPRGSGWPLGEGITGYLKYPTATYATLGDPMQ	70	
DB	12	PMSEITLIIYSISFFEGGLSLRLRR--RLYPPGPKLPLVG--SMHMMQDITTHGLKIAKKO	68	
QY	71	HVSKGKTYRSLJFEPTIVSADAGLNFIIONEGRLFE-----CSRPRSIGTL	120	
DB	69	-----YGGLEFHRMKGILHWYTVSSPEIARQVQVONITSNRPANAIISTYIYTRAD----	120	
QY	121	GRMSMLVLVGMHHRDMRSIS--LNFLSHARLT-----ILKDNVERH-----T	161	
DB	121	---MAFAHYGCFWROMRKICVAKLFESRRKRAESWESVRDEVDLSMLKTVEANICKPVNLGEL	177	
QY	162	LEVLDWMQONSIFESQDRAKKTFFMLMKHKHMSMDPGBEETQOLKREYTFPKGYVSAPL	221	
DB	178	IFTL--TMNITTYRAFAK-----NEGODEFTKILQER--SKLGFAPNMSD	219	
QY	222	NLPCTAYARAKAOSRATILIKFTIERKMEERKLD--IKEEQEIEEVTEDEAEKMSKD----	275	
DB	220	FIPMIGWIDPGGLSRLVK--ARKALDFIDSIIIDHILQKRQNNFSDAEIETWDDMLA	277	
QY	276	-----HVKQRIIDDLLGWLKHSNISTQIIDLILSLPAGHEISSVAIALAIFPLQACP	331	
DB	278	FYGEARARVDESDDL-----QKAISLTKNIKAIIMDVNFGGTEYVAASILEWMAELMKSP	333	
QY	332	KAVELEIREHEEIAARAKKEGSELNMDQYKMDPTQCVINTEITLGNVRFIRHAKALKD	391	
DB	334	EDQKRVQDELAVVGLERVERESDID-----KLTFPLKALKETTLKNNHPIPLLLHETSQD	388	

```

OY 392 VRYGYDIPSGWKVLPVISAVALHNSRYDQPNLFNPMWMOOONNGASSGSGSFSTWGN 451
DB 389 AEVAGYIFPKQTRVMAVAIGRDKNSMEDPDAFKPSRLK-----PGVDFK--GNH 439
OY 452 --YMPFGGPRLCAGSELAKLEMAVFIHLVLKFMWELAEEDQPAFPFVFPNGPLPIRV 509
DB 440 FEPIFGSGRRSCPGMQLGTYITLDIAVAHLHLCFTWELPDGKMPSELMDMFGILTAPRA 499
OY 510 SRIT 513
DB 500 TRIV 503

```

RESULT 4

```

US-09-796-256A-4
; Sequence 4, Application US/09796256A
; Patent No. US20020078477A1
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent L
; APPLICANT: Cattray, Daniel R
; APPLICANT: Smeltzer, Richard H
; TITLE OF INVENTION: Production of Stryngyl Lignin in Gymnosperms
; FILE REFERENCE: 50617/G-3532.0
; CURRENT APPLICATION NUMBER: US/09/796,256A
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US60/033381
; PRIOR FILING DATE: 1996-12-16
; PRIOR APPLICATION NUMBER: 08/991677
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Liquidambar styraciflua
US-09-796-256A-4

```

```

Query Match
Best Local Similarity 10.8%; Score 289; DB 10; Length 511;
Matches 121; Conservative 95; Mismatches 196; Indels 138; Gaps 21;

```

```

OY 6 HHTLLPL-LPLPSLLFLILLKRNKTRNLPKSGMPLGEGTYLKYPTATTL 64
DB 6 HEALQPLPMTLFFILLGLVSR--LQRLPYPPGPKGLPIGNML-MMDOLTHGCL 62
OY 65 GDFMOHVSXGKIYNSNLFGEPTIVSADAGNRFILONEGRLEF-----CSTPR 114
DB 63 AKLAKO-----YGGLEHLKMGFLHMAVSTPDMAQVLOYODNIFSRRPATTIAISYLYDR 118
OY 115 SIGGILGKMSMLVLDGMDRMRKRSIS-LNFISSHARLRTILLKDVERRHTLFVLDGMOQNSI 173
DB 119 AD-----MAFAHYGPFWRQMRKLCYMKLSRRK-----AESWE----- 151
OY 174 FSAODE-----AKKFTENLMKHHMSMDPE-----EETE 203
DB 152 -SVRDEVDAAVAVASNGSTVNIIGELVFALTKNITRYBAFGTISHDDODEVAIIQEFST 210
OY 204 QI-----KKEYVTENKGV--VSAPLNLPGTAYKALQSAATILKTERMEERKIDIRE 255
DB 211 QIPLGAFNIADIFPLMKWPOGINVRNL-----KARGALDGHIDKIDH---IQK 257
OY 236 EDOEEEVKTE-----DEAKMSQDVHRRKORTDIDLGLWVFKHNSNLTBOIDLIT 305
DB 258 GSKNSEEDVDMDLAFYGEELKVSSED-----LQNSIKLTKONIKAI 303
OY 306 LSLFAGHERSSVAIALAIFFLQACRAVEBELREHLEIARAKKELGSESELNMDYKMD 365
DB 304 MDVAFGGSTEVASAIEMAMTELKSPEDLAKYQOELAAVVGIDRVEREK-----DFEKLTP 358
OY 366 FTQCVINETLRIGNVVRFHLRAKLDVARKYGDIPSGWKVLPVISAVALHNSRYDQPNLF 425
DB 359 YLKCIVKEVLRHPPIPLLLHETAEADAEVGGYIIPAKSRVWINACALGRKNSHADDPTF 418

```

```

OY 426 NPMWMOOONNGASSGSGSFSTWGN--YMPFGGPRLCAGSELAKLEMAVFIHLVLKRF 483
DB 419 RSERFLK-----DGVDFK--GNHFEPIFGSGRRSCPGMQLGTYITLDIAVAHLHLCF 469
OY 484 MWELAEEDQ 493
DB 470 TWELPDGMRP 479

```

RESULT 5

```

US-09-739-254-69
; Sequence 69, Application US/09739254
; Patent No. US20010021700A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032p1
; CURRENT APPLICATION NUMBER: US/09/739,254
; PRIOR FILING DATE: 2000-12-19
; EARLIER APPLICATION NUMBER: 09/511,554
; EARLIER FILING DATE: 2000-02-23
; EARLIER APPLICATION NUMBER: PCT/US99/19330
; EARLIER FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 60/097,917
; EARLIER FILING DATE: 1998-08-25
; EARLIER APPLICATION NUMBER: 60/098,634
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 69
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (502)
; OTHER INFORMATION: Xaa equals stop translation
US-09-739-254-69

```

```

Query Match
Best Local Similarity 10.6%; Score 285.5; DB 10; Length 502;
Matches 124; Conservative 86; Mismatches 210; Indels 109; Gaps 22;

```

```

OY 15 LPSLLSLFLFL-----LTKRNKTRNLPKSGMPLGEGTYLKYPTATTLID----- 66
DB 14 LGGALFLFLFLGLVROLKOR--PMGFPPGPGPLFIGNI-----YSLAASSELPHV 64
OY 67 PMOQHVSKYGIYNSNLFGEPTIVSADAGNRFILONEGRLEFCSY-----RSIGG 118
DB 65 YMRQSOVYGEISLIDGGISYVANGDYVKECLVHSELP-ADRPCLPLFMKWTMGW 123
OY 119 ILGR-----MSMLVLDGMDRMRKRSISLNFISSHARLRTILLKDVERRHTLFVLDGMOQNSI 173
DB 124 LLSNRYGRGV-----DHRRLAANSFRYFGYGO-----KSPESKIL----- 164
OY 174 FSAODEKK--PTPMMAKHHTMS-----MDPE-----EETE-----QIKKEYVTENKGV 216
DB 165 FNAITTYIGRPDRKQILTNAVSNITMLIFGERFTYEDIDPQHMTLESENVELAASA 224
OY 217 VSAPLN-----LPQTAVHKAQSRATILKTERKKEERKIDIREDOEEEVKTEDEA 269
DB 225 SYELNAPFWIGILFPGHQDLFRNAAVYDFLSRLTEKASVNRKPOLQHFVDATLDEM 284
OY 270 ESKSDHVRKORTDDDLGLWVFKHNSNLTBOIDLITLSLFAGHERSSVAIALAIFFLQ 329
DB 285 DQGNKD-----PSTFSKENLIFSGELINAGTETTTNVLMWALIFMAL 328
OY 330 CPKAVEELREHLEIARAKKELGSESELNMDYKMDPQCVINETLRIGNVVRF-LHRA 388
DB 329 YPNIGGOVQRE-IDLIMGF-----NGKPSMDCKKMPYEAVALHEVLRFCNIVPLGIFHAT 383
OY 389 LKDVRYGYDIPSGWKVLPVISAVALHNSRYDQPNLFNPMWMOOONNGASSGSGSFSTW 448
DB 389 YLKCIVKEVLRHPPIPLLLHETAEADAEVGGYIIPAKSRVWINACALGRKNSHADDPTF 418

```

Db 384 SEDAVVRSYIPKGTIVITNLVSVHDEKYMREDFVHBERF-----LDSSGYFAK 434
QY 449 GNNYMPFGGPRLCAGSELAKLEMAVFIHHLV-----LKENWELAEEDDP 493
Db 435 KEALVPSISGRRCGCEHLARMEMFLFTALLORFHLHPHVELVPLDKP 483

RESULT 6

US-09-904-615-69
; Sequence 69, Application US/09904615
; Patent No. US2002026040A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/904,615
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 69
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (502)
; OTHER INFORMATION: xaa equals stop translation
US-09-904-615-69

Query Match 10.6%; Score 285.5; DB 10; Length 502;
Best Local Similarity 23.4%; Pred. No. 2,6e-15;
Matches 124; Conservative 86; Mismatches 210; Indels 109; Gaps 22;

QY 15 LPSLSLLFLI---LKRNRKTRNLPGKSGMPLEGITGYLKPYTATTLGD---- 66
Db 14 LGALFLLFLFALGVRLQKRR---PMGPPGPPGLPIGNI-----YSLAASSELPHY 64
QY 67 ENQOHKSKYKGIYRSLFGEPTIVSADAGLNFILQNEBRLFECSYR-----RSTGG 118
Db 65 YNRKSOQVGEIFSLDGGISTVVLNGYDVKECLVHSEIR-ADRPCPLRFMKMTKMG 123
QY 119 ILGK-----WSMLVLVGDHGRDMRSISLNFSLHARLRTILKDVERTLEVLDSMOQNSI 173
Db 124 LNSRGRGW-----DHRRLVNSFRYPGCG-----KSFESKIL-----ETRF 164
QY 174 FSAQDEAKK---FTENLMAKHIMS-----MDPGE---ETE---OLKKEVYFMKGV 216
Db 165 FMDALETYKGRFEDFKQLITNAVSNTNLIIFGERFETEDTOFOHMIELFSEVELASA 224
QY 217 VSAPLN-----LGTANHALOSRATILKFIKMERKADIKEDDEEEVYKTEDA 269
Db 225 SVFLYNAFPGICILPPGKHQQLFRNAAYVDLFLIEKASVNRKQPLQHFNVDAYLDEM 284
QY 270 EKSQSDHVKORTDDLLGLWVLKHSNLSTEQIIDLILSLFGHETSSVALAIFELQA 329
Db 285 DQGNKD-----PSSTFSKENLIFSVGELLINGTETTTVNLVALLIFMAL 328
QY 330 CPKAVEELREHLEIARAKKELGESELMNDYKKMDFTQCIVINETLRIGNVVR-LHRRK 388
Db 329 YENIGOVQKE-IDLIMGP-----NGKPSWDCKCKMPYTEAVLHEVLRFECNIVPLGIFHAT 383
QY 389 LKDVATKGDIVSSGWKVLPIYSAVHLDSNRYQPNLFNWRMOQONNGASSGSSGFSFTW 448
Db 384 SEDAVVRSYIPKGTIVITNLVSVHDEKYMREDFVHBERF-----LDSSGYFAK 434
QY 449 GNNYMPFGGPRLCAGSELAKLEMAVFIHHLV-----LKENWELAEEDDP 493

Db 435 KEALVPSISGRRCGCEHLARMEMFLFTALLORFHLHPHVELVPLDKP 483

RESULT 7

US-10-067-668-10
; Sequence 10, Application US/10067668
; Publication No. US20030022334A1
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria Alexandra
; TITLE OF INVENTION: 33312, 33303, 32579, NOVEL HUMAN
; FILE REFERENCE: 10448-136001
; CURRENT APPLICATION NUMBER: US/10/067,668
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/266,140
; PRIOR FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 10
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-10-067-668-10

Query Match 10.3%; Score 275; DB 9; Length 496;
Best Local Similarity 23.6%; Pred. No. 1.8e-14;
Matches 126; Conservative 82; Mismatches 208; Indels 118; Gaps 17;

QY 40 PEGKSGMPLEGITGYLK-----PYTATTLGDPMQOHVSKYK-ITRSLFGEPTIVSAD 93
Db 1 PEGPPLPLIGNLQLGRAPGPIPHSLTKL-----RKARVYGPVETLVIGRPVVVLG 55
QY 94 AGLNRFILONEGR-----FECYSRISIGILGKSMVLVGDHGRDMRSISLNF-- 144
Db 56 PEAVKRVLLDKGEERFAGKGDENFTPTPLSKYREGGLFSDNGPRWKLRRSLTLTF 115
QY 145 -----SHARLRTILKDVERTLEVLDSMOQNS-IFSADEAKKFTFMLAKHNS 194
Db 116 HFGMAYSKRSQLEPRIOEARLDYERLRKEQAGSPIDITELLARLPLNVICSLG 175
QY 195 MDGGEETQLKKEVYTFPMKGVSAPLN-----LGTANHAL 232
Db 176 V-----RFYLRPEDEFLK-LIDKLNMFRVSPWQOLDIFPRLRYLPGLSFRKAF 229
QY 233 GSRATILKFIKMER-----RKLDIKEEQEEVEVYKTEDAEMSKSDHYKQR 281
Db 230 KAAKLDKLDLIERRETLERPADPRDIDGFLDSLLEKKBGANKSE----- 281
QY 282 TDDDLGLWVLKHSNLSTEQIIDLILSLFAGHETSSVALAIFELQACPKAVEELREH 341
Db 282 -----LSDEELAAVTVLDLIFAGTETTSSTLSMALYLLAKHREVOAKLREI 327
QY 342 LEIARAKKELGESELMNDYKKMDFTQCIVINETLRIGNV-RELIRKALKDVRV-KGIYI 399
Db 328 DEVIGDR---SPYDVDAQRAQMPYIDAVIKETLRLYPVPLLPVAKKDEIPGUYI 384
QY 400 PSGWKVLPYISAVHLDSNRYQPNLFNWRMOQONNGASSGSSGFSFTW 459
Db 385 PKGTIVYLVYLNHDPKFFPNPEEDPERFLDENKFKKSTA-----FLPGAGP 435
QY 460 RLCAGSELAKLEMAVFIHHLVLFKFWWELAEEDDPAFAPEVDEPN-----GLP 506
Db 436 RNCIGERLARMBELFLFATLLORF-----PLELAIVPGDIPSLTPKPELGIP 483

RESULT 8

US-09-945-301-4
; Sequence 4, Application US/09945301
; Patent No. US200209099A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.

; APPLICANT: Glucksmann, Maria
 ; APPLICANT: Tsai, Fong-Ying
 ; TITLE OF INVENTION: 27439, NOVEL HUMAN HYDROXYLASE AND USES
 ; TITLE OF INVENTION: THEREFOR
 ; FILE REFERENCE: 38155-20036.00
 ; CURRENT APPLICATION NUMBER: US/09/945,301
 ; CURRENT FILING DATE: 2001-08-31
 ; PRIOR APPLICATION NUMBER: US 60/229,301
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 496
 ; TYPE: PRF
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Consensus amino acid sequence
 ; US-09-945-301-4

Query Match 10.3%; Score 275; DB 10; Length 496;
 Best Local Similarity 23.6%; Pred. No. 1,8e-14;
 Matches 126; Conservative 82; Mismatches 208; Indels 118; Gaps 17;

QY 40 PRGKSWPPLGRTIGYK-----PYATITLGDGMQOHVSKYK-ITRSNLFGEPTVSAD 93
 DB 1 PRGPPPLPLIGNLQGRAPGPIPHSLTKL-----RKAKYGRPVFTLYLGPVVVLTG 55
 QY 94 AGINRFLINEGR-----FECSTPRSIGILGKMSMLVYGDHMRDRSISLNF-- 144
 DB 56 PEAVKEVLIDKGEFPAKGDENFTFPMLSKGRREGILFSDGPKRKIRRSLLTIR 115
 QY 145 -----SHARLTILKDVRRHTLVLDVSDQONS-IFSADCAKKTFTNLMARHMS 194
 DB 116 HFGMGAYSRKSOCLPEPRIOEFARDLYERLRKQAGSPIDITELARLAPLAVYCSILFG 175
 QY 195 MDGGEETQTKKEYTTPMKGVYAPLN-----LPGTAYHKAL 232
 DB 176 V-----RFDLRPEDEPEFLK-LIDKILNEMEDRVSPMHOLLDFPPLLRYLPGLSKRAF 229
 QY 233 OSRATILKFERKME-----RKIDKEEDOEERETEDAEKMSKDHVKQR 281
 DB 230 KAKAKLDKIDKLEERRETEPAGDPRRLDIFGLSLLLEAKREGNPKSE----- 281
 QY 282 TDDDLGWLKHSNSTEQIILILSLFAGHETSSVAIALAIFLOACRAVEELREH 341
 DB 282 -----LSDEELATVLDLTFAGTETSTLSMALYLLAKHPEVQAKLEEI 327
 QY 342 LETARAKKEGESELMWDYKMDFTQCVINETLRGNY-RELRKALDVAK-KGYDI 399
 DB 328 DEYIGDR--SPTTYVDARAQMPYDAVYKETLRYPVPLLPVATVDTETIPDGYL 384
 QY 400 PSGMKVLPVISAVALHNSRYDQPNLFNPMWMOOONNGASSSGSFGSTGNMMPFGCGP 459
 DB 385 PKGTIVYVNLISLRDPKYPNPEEPDEFKLDENGKFKKSYA-----FLPFGAGP 435
 QY 460 RLCAGSELAKLEMAVFIHLVLKFMNELAEDDOPFAFPVDFPN-----GLP 506
 DB 436 RNCIGERLAMELFLATLQRF-----PELELAVPPDIPSLPVPKPELGLP 483

RESULT 9
 ; US-10-067-668-12
 ; Sequence 12, Application US/10067668
 ; Publication No. US20030022334A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Glucksmann, Maria, Alexandra
 ; TITLE OF INVENTION: 33312, 33303, 32579, NOVEL HUMAN
 ; FILE REFERENCE: 10448-136001
 ; CURRENT APPLICATION NUMBER: US/10/067,668
 ; CURRENT FILING DATE: 2002-02-04
 ; PRIOR APPLICATION NUMBER: 60/266,140
 ; PRIOR FILING DATE: 2001-02-02

; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12
 ; LENGTH: 470
 ; TYPE: PRF
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Consensus sequence
 ; US-10-067-668-12

Query Match 10.1%; Score 272; DB 9; Length 470;
 Best Local Similarity 28.6%; Pred. No. 3e-14;
 Matches 87; Conservative 47; Mismatches 110; Indels 60; Gaps 8;

QY 223 LPGTAYHKALOSRATILKFERKME-----RKIDKEEDOEERETEDAEK 271
 DB 194 LPGLSKRAKAKAKDKLDKIDKLEERRETEPAGDPRRLDIFGLSLLLEAKREGNPK 253
 QY 272 SKSDHVRKQRTDDDLGWLKHSNSTEQIILILSLFAGHETSSVAIALAIFLOACP 331
 DB 254 SE-----LSDEELATVLDLTFAGTETSTLSMALYLLAKHP 291
 QY 332 KAVEELREHLEIARAKKEGESELMWDYKMDFTQCVINETLRGNY-RELRKAL 390
 DB 292 EVQAKLEEDIDEYIGDR--SPTTYVDARAQMPYDAVYKETLRYPVPLLPVATK 348
 QY 391 DVRY-KGYDIPSGMKVLPVISAVALHNSRYDQPNLFNPMWMOOONNGASSSGSFGTWG 449
 DB 349 DTEIPDGYLIPKGTIVYVNLISLRDPKYPNPEEPDEFKLDENGKFKKSYA----- 401
 QY 450 NNTMPEGGPRLCAGSELAKLEMAVFIHLVLKFMNELAEDDOPFAFPVDFPN----- 503
 DB 402 --FLPFGAGPRNCLGERLAMELFLATLQRF-----PELELAVPPDIPSLPVPKPE 453
 QY 504 GLP 506
 DB 454 LGLP 457

RESULT 10
 ; US-09-992-901-2
 ; Sequence 2, Application US/09992901
 ; Patent No. US20020073446A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Neff, Michael M.
 ; APPLICANT: Cho, Joanne
 ; TITLE OF INVENTION: GENETICALLY MODIFIED PLANTS HAVING
 ; FILE REFERENCE: SALKINS.024DV1
 ; CURRENT APPLICATION NUMBER: US/09/992,901
 ; CURRENT FILING DATE: 2001-11-14
 ; PRIOR APPLICATION NUMBER: US 60/124570
 ; PRIOR FILING DATE: 1999-03-16
 ; PRIOR APPLICATION NUMBER: US 60/170,931
 ; PRIOR FILING DATE: 1999-12-14
 ; PRIOR APPLICATION NUMBER: US 60/172,832
 ; PRIOR FILING DATE: 1999-12-20
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 520
 ; TYPE: PRF
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-992-901-2

Query Match 9.6%; Score 256.5; DB 10; Length 520;
 Best Local Similarity 23.8%; Pred. No. 6.3e-13;
 Matches 136; Conservative 95; Mismatches 225; Indels 115; Gaps 25;

```

Db 1 MEESSMFIPIVVL--VLSTVLSTVYVGMILMKRPKIEBHNSKQIGRPPIFFTG 57
Oy 51 ---ETIGVL-----KPYATTLGDFMOOHVSKYKIRSNL--FGPPT--IYADAG 95
Db 58 NKEVLGVMILKASHPMPHSNILLRVLSEY--HHMKIYGAFVLWFG--PTFRLIYVADPD 115
Oy 96 LNRFLIÖNGRLEFECSPRSIGIGILGKMSMLVVGDM---HRDMSISLNFSLHARLRI 152
Db 116 LIRETF-SKSEFYEKNKNAHPLVKOLEGDLISLKGKMAHHRKILISPTHEMELKLLVYV 174
Oy 153 LKQVDRHTLFLVLSWQ-----ONSTFSADQEAKKPTFN 187
Db 175 VLKSTVD---AVDKRSDKLSNGEYEVYVEMFOITLEDVLSRTAFSGSYEDGRVFL 230
Oy 188 MAKHIMSDPGEETEQLKREYVETMKGVASPL--NLPGTAYHKAQSRATILKEIK 245
Db 231 QAQOMILC-----AEAFQK---VFIPGYRFPPTGRLNLSRKLDKEI--RKSILKIEHR 279
Oy 246 MEERKLDIKEEDQEEVEVYVTEDEAKESKSDHYRKQRTDGLGVLKHSNLSSTEQILDLI 305
Db 280 -ROMAIDGEGECKEPAAK-----DLGLMLQANVNVODIVEEC 318
Oy 306 LSLFAGHETSSVAIALAIFFLQACPRAVEELREHELEIARAKKEGSELMWDDY--KKM 364
Db 319 KSEFFRAGKOTISNLLTWITLISMPHEMOAKARDEVLRVC-----GSDVPTPKDHVYVL 372
Oy 365 DETQCVINETRLGNVRFRLKALKDVRYKGYDIPSGMKVLPVISAHLNDSRY--DQPN 423
Db 373 KTLMSILNESLRLPPYIVATIRRAKSDVLYGKIKIPCGTELLIPILAHHQDAIMQNDVN 432
Oy 424 LFNPRMOQNNNGASSSGSFGSTGNNMPPGCGPRCLGSELAKLEMAVFIHHLVLYKF 483
Db 433 ENPAPRAFADGVPRAKHPVG-----FIPGLGVKTCIGONLAILQAKTLAVMIQRF 484
Oy 484 NMEIAEDQ--PPAFPVDFPGLDIRVSR 512
Db 485 TFLAPTYGHAPTYMLLTPGHGAPITFRRL 515

RESULT 11
US-09-796-138-19
; Sequence 19, Application US/09796138
; Patent No. US20020031782A1
; GENERAL INFORMATION:
; APPLICANT: Waterman, Michael R.
; APPLICANT: Bellamine, Aouatef
; APPLICANT: Podust, Larissa M.
; TITLE OF INVENTION: Mycobacterium tuberculosis cyp51 HIGH RESOLUTION STRUCTURE, POLYPE
; TITLE OF INVENTION: AND
; TITLE OF INVENTION: NUCLEIC ACIDS, AND THERAPEUTIC AND SCREENING METHODS
; TITLE OF INVENTION: RELATING TO SAME
; FILE REFERENCE: Attorney Docket No. US20020031782A1 1242-17-2
; CURRENT APPLICATION NUMBER: US/09/796,138
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/345,218
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Penicillium italicum
US-09-796-138-19

Query Match 9.4%; Score 251; DB 10; Length 515;
Best Local Similarity 22.7%; Pred. No. 1.8e-12;
Matches 120; Conservative 75; Mismatches 213; Indels 120; Gaps 21;
Oy 12 LLLPSSLALLFLILKRRNRKTRFNLPPKSGM--PFLGRTIGY-LKPYTATTLGDFMQ 69
Db 21 LFLVSLVNLVNIQILFYNRKE-----PPVYFHWIPIFGSTIAYGMDPY-----QFFF 67
Oy 70 QHVSRYGKIYRSNLFGEPTIYVADAGLNRFILÖNGRLEFECSPRSIGIGILGKMSMLVYV 129

```

```

Db 68 ASRAKYGDIFTFILLOKKTIVYLGVGNEFIL--NGKLKDVAE-----VYGRITTPVFG 121
Oy 130 GDMHRMRISL-----NPLSHARLFTILKDVDRHLLFVLDSMOQNSIFSADQEAK----- 181
Db 122 SDVYVDCPSKMLEQKFKYIGLSQAL-----ESTYPLAD--ETNAYIKSSPNEFGQSG 175
Oy 182 -----KTFENMAKHIMSDPGEETEQLKREY-----VPEMKGVSA 219
Db 176 TIDLAAMAIEITFTAARLIQ-----GEEVRSKLTSEFADLFHDLDGFSPIINMLPWAPL 231
Oy 220 PLN---LPGTAYHKAQSRATILKFERKMEKLDIKEEDQEEVEYKTEDEAKESKSDH 276
Db 232 PHNAAIKHTTYARDLSGNVPASATGSMRRQRROD----- 267
Oy 277 VRKQRTD--DDLGVYLKHSN--LSTEQILDLILSLFAGHETSSVAIALAIFFLQACPRA 333
Db 268 -KSKGTDMSNLKRCYRGRGTPRPDKEIAHMTTLMAQOHSSAISCIILLRLASQPEK 326
Oy 334 VEELEHELEIARAKKEIGES--ELNWDYKKMDFQCVINETRLGNVRFRLHRRALD 391
Db 327 AERLHAEQI-----KNLGADLPPLQYKDMDKLPLLRNVIKETLRHSSIHFLMRKVKNP 380
Oy 392 VRYKGYD--IPSGMKVLPVISAHLNDSRYDQPNLEPNRMOQ-----NNGA 437
Db 381 MPVPGTDFVPPSHSLSSPGVTARDEHFRDPLRMDPHRWESRYVEDSSDTVDYGYGA 440
Oy 438 SSSGSGSFGSTGNNMPPGCGPRCLGSELAKLEMAVFIHHLVLYKF 485
Db 441 VSKGTS-----PYLPFGAGRHCIGEFAYLNTLEVYATVLYRERF 482

RESULT 12
US-09-909-903-19
; Sequence 19, Application US/09909903
; Patent No. US20020052031A1
; GENERAL INFORMATION:
; APPLICANT: Waterman, Michael R.
; APPLICANT: Bellamine, Aouatef
; APPLICANT: Mycobacterium tuberculosis cyp51 POLYPEPTIDES AND
; TITLE OF INVENTION: NUCLEIC ACIDS AND THERAPEUTIC AND SCREENING METHODS
; TITLE OF INVENTION: RELATING TO SAME
; FILE REFERENCE: Attorney Docket No. US20020052031A1 1242-17
; CURRENT APPLICATION NUMBER: US/09/909,903
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Penicillium italicum
US-09-909-903-19

Query Match 9.4%; Score 251; DB 10; Length 515;
Best Local Similarity 22.7%; Pred. No. 1.8e-12;
Matches 120; Conservative 75; Mismatches 213; Indels 120; Gaps 21;
Oy 12 LLLPSSLALLFLILKRRNRKTRFNLPPKSGM--PFLGRTIGY-LKPYTATTLGDFMQ 69
Db 21 LFLVSLVNLVNIQILFYNRKE-----PPVYFHWIPIFGSTIAYGMDPY-----QFFF 67
Oy 70 QHVSRYGKIYRSNLFGEPTIYVADAGLNRFILÖNGRLEFECSPRSIGIGILGKMSMLVYV 129
Db 68 ASRAKYGDIFTFILLOKKTIVYLGVGNEFIL--NGKLKDVAE-----VYGRITTPVFG 121
Oy 130 GDMHRMRISL-----NPLSHARLFTILKDVDRHLLFVLDSMOQNSIFSADQEAK----- 181
Db 122 SDVYVDCPSKMLEQKFKYIGLSQAL-----ESTYPLAD--ETNAYIKSSPNEFGQSG 175
Oy 182 -----KTFENMAKHIMSDPGEETEQLKREY-----VPEMKGVSA 219
Db 176 TIDLAAMAIEITFTAARLIQ-----GEEVRSKLTSEFADLFHDLDGFSPIINMLPWAPL 231
Oy 220 PLN---LPGTAYHKAQSRATILKFERKMEKLDIKEEDQEEVEYKTEDEAKESKSDH 276
Db 232 PHNAAIKHTTYARDLSGNVPASATGSMRRQRROD----- 267
Oy 277 VRKQRTD--DDLGVYLKHSN--LSTEQILDLILSLFAGHETSSVAIALAIFFLQACPRA 333
Db 268 -KSKGTDMSNLKRCYRGRGTPRPDKEIAHMTTLMAQOHSSAISCIILLRLASQPEK 326
Oy 334 VEELEHELEIARAKKEIGES--ELNWDYKKMDFQCVINETRLGNVRFRLHRRALD 391
Db 327 AERLHAEQI-----KNLGADLPPLQYKDMDKLPLLRNVIKETLRHSSIHFLMRKVKNP 380
Oy 392 VRYKGYD--IPSGMKVLPVISAHLNDSRYDQPNLEPNRMOQ-----NNGA 437
Db 381 MPVPGTDFVPPSHSLSSPGVTARDEHFRDPLRMDPHRWESRYVEDSSDTVDYGYGA 440
Oy 438 SSSGSGSFGSTGNNMPPGCGPRCLGSELAKLEMAVFIHHLVLYKF 485
Db 441 VSKGTS-----PYLPFGAGRHCIGEFAYLNTLEVYATVLYRERF 482

```

OY 220 PLN---LPGTAVHKAQSAFTIKETERMEERKLDIKEEDOEFEVKTEDENKSDH 276
 DB 232 PHNASHAKHTFYARDLSGVNPSATGSMRRORRRD----- 267
 OY 277 VKORPDD--DILLGWLJKSN--LSTEOJLDLITSLFAGHETSVAIALAIFLOACPKA 333
 DB 268 -KSKGDMISMLRCVYRGTPIPDKEIAHMMITTLMAQOHSASISMLLASOPEK 326
 OY 334 VEREHELEIARAKKELES--ELNMDYKKMDFTQCIINETRLGNVVRFLRRKALD 391
 DB 327 AEKLAHBOI-----KNIGADLRLQYKMDKPLRLANVYKELRLHSSHTLMRKVKNP 380
 OY 392 VRYKYD--IPSGWVLPVISAHVLDNSRYDOPNLFNPMWQOQ-----NGA 437
 DB 381 MPYGTDFVVPSPHTLSSPGVTARDEHRDPLRMWDPHRESRVEVEDSDPTVGYGA 440
 OY 438 SSSGSGSFTWGNMYPFGGPRLCAGSELAKLEMAVFHHLVLRKNW 485
 DB 441 VSKGTRS-----PYLPFGAGRHCIGEFAYALNLEVIATVREFR 482

RESULT 13

US-10-067-668-8
 ; Sequence 8, Application US/10067668
 ; Publication No. US20030022334A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Glucksmann, Maria Alexandra
 ; TITLE OF INVENTION: 33312, 33303, 32579, NOVEL HUMAN
 ; FILE REFERENCE: 10448-136001
 ; CURRENT APPLICATION NUMBER: US/10/067, 668
 ; CURRENT FILING DATE: 2002-02-04
 ; PRIOR APPLICATION NUMBER: 60/266, 140
 ; PRIOR FILING DATE: 2001-02-02
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 544
 ; TYPE: PRP
 ; ORGANISM: Homo sapiens
 US-10-067-668-8

Query Match 9.4%; Score 251; DB 9; Length 544;
 Best Local Similarity 23.2%; Pred. No. 1.9e-12;
 Matches 127; Conservative 98; Mismatches 234; Indels 88; Gaps 23;

OY 13 LILPLSLILFLILKRRNRKTRFNLPGKSGMPFLGRTGY--LKP----- 59
 DB 37 LILGLVALLGWSMLRRRRAR---GIPGPTWPLVG--NFGVLLPFLRRSSWLSRT 91
 OY 60 ----TATTLG-DPMOQHVSK-YGKIYRSMJGEPPIVSADAGLNRFILQNEGRLEPCSY 112
 DB 92 RAAGIDPSVIGPVLALHARVGSIFSGFIHVLVYVLSDFHSVRALVQAQAEVFS-DR 150
 OY 113 PR-SIGGILGKMSLVYV--GDMHRDMSISLNFSLHARLTILL--KDVRRHTLEVYDS 167
 DB 151 PRVPLISIVYKKEGVFAHGPVWROQRKFSHTLRHFGKLSLEPKIIEERK-YVKA 209
 OY 168 MOONSIFSADDEAKKTF--NIMAKHIMSMDPE--EETQOLKEVYTFKGVASAPLN- 222
 DB 210 MOKH---GDDPCPSFTISNAVSNITSLCFQGRDYTSSEFKMKGMSRGLEICLNS 265
 OY 223 -----LPGTAVHKAQSRATILFKTERMEERKLDIKEEDOEFEVKTEDEN 269
 DB 266 QVLLVNICPMLVYLPFGPFEKELRQIEKDITSL-----KTIKQGESLDRENPDPI 318
 OY 270 ESKSDHVRKORTDDLLGVNKLHNSLSTEOJLDLITSLFAGHETSVAIALAIFLOA 329
 DB 319 DMYLL-HMEERKNN-----SNSSFEDEYLYFIIGDLEIAGDTTNSILMCLYMSL 370
 OY 330 CPRAVEELREHELEIARAKKELESSELMNDYKKMDFTQCIINETRLGNVVR-LLRKA 388
 DB 371 NPVOQEVHEIEIRVIGANR-----APSLDKAKMPYTEATIMEVORLTJVVPPLAIPHT 425

OY 389 LKDVRYKYDIPSGWVLPVISAHVLDNSRYDOPNLFNPMWQOQNNGSSGSGSFTW 448
 DB 426 SENTVLQGYTIPKGTLLILNLSVHRDPALWEKPEDFYFNRLFDGGLIKK----- 477
 OY 449 GNNYMPFGGPRLCAGSELAKLEMAVFHHLVLRKWELEAD-OPF---APPYDFENG 504
 DB 478 -ETFIFFGIGKRVQGEOLAKWELFLMFVSLMQSFALPDESKPILTRGRLTLAPHP 536
 OY 505 LPIVRS 511
 DB 537 FNITISR 543

RESULT 14

US-09-796-138-18
 ; Sequence 18, Application US/09796138
 ; Patent No. US20020031782A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Waterman, Michael R.
 ; APPLICANT: Bellamine, Aouatef
 ; APPLICANT: Podust, Larissa M.
 ; TITLE OF INVENTION: Mycobacterium tuberculosis cyp51 HIGH RESOLUTION STRUCTURE, POL
 ; TITLE OF INVENTION: AND
 ; TITLE OF INVENTION: NUCLEIC ACIDS, AND THERAPEUTIC AND SCREENING METHODS
 ; FILE REFERENCE: Attorney Docket No. US20020031782A1 1242-17-2
 ; CURRENT APPLICATION NUMBER: US/09/796,138
 ; PRIOR APPLICATION NUMBER: 09/345, 218
 ; PRIOR FILING DATE: 1999-06-30
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 18
 ; LENGTH: 503
 ; TYPE: PRP
 ; ORGANISM: Homo sapiens
 US-09-796-138-18

Query Match 9.3%; Score 249; DB 10; Length 503;
 Best Local Similarity 24.1%; Pred. No. 2.5e-12;
 Matches 132; Conservative 87; Mismatches 212; Indels 116; Gaps 28;

OY 9 LILPLLLPLSLD-LFLILKRRNRKTRPNLPK-----SGMFLGRTGYLKYPT 60
 DB 26 LLSMLIACAFITSLVLYLRLAAGH--LVQLPAGVSPYIRSPRFLGHALAFKSP 82
 OY 61 ATTIGDPMOQHVSKYKTIYRSMJGEPPIVSADAGLNRFILQNEGRLEPCSYPSIG 118
 DB 83 ----EFLNAYEKYGVFSFTVVGKTFYILGSDAALLFNSKNEDLNAMEDYYSRLTTP 137
 OY 119 ILGKMSLVYVGDW---HRDMRSISLN--FLSHARLTILLKDVRRHTLEVYDSMOON- 171
 DB 138 VEGKGVAVDYVPNPFLEOKMKLSGLNIAHFKQVSI-----IEKETYEYESMGESG 190
 OY 172 -SIFSADDEAKKTFNIMAKHIMSMDPGEETEOLKE---YTFMKGVASAPLNLP 225
 DB 191 EKNVEALSE---LILTFASHCHL--GKEIRSQLNEKVAOLYADDDGFSHAMVLLP 243
 OY 226 ----TAVHKAQSRATILFKTERMEERKLDIKEEDOEFEVKTEDENBMSKSHVAKOR 281
 DB 244 WLPSPFRDRRAHREIKDIFYRAIOKRR-----OSOKI----- 278
 OY 282 TDDDLGLVNLKHS-----NLSTEOJLDLITSLFAGHETSVAIALAIFLOACPKAVEE 336
 DB 279 --DDILOTLDATYKORPLTDEVAGMLGILLAGHTSSTISAMWGFTL-ARDKTLQ- 334
 OY 337 LREHELEIARAKKELES--ELNMDYKKMDFTQCIINETRLGNVVRFLRRKALKDVR 394
 DB 335 -KKCYLE--QKVCGENLPLVYDQKDLNLDRCIKETRLRLRPPIIMIMMARTRPQTV 390
 OY 395 KGDIPIGKMYL--PVISAHVLDN--SRYPNLFNPMWQOQNNGSSGSGSFTWGN 450

Db 391 AGTTPPGHVCSPVNOBKDSWVERLD-----FNPDRILODN---PASGE-KFA----- 438
 QY 451 NYMPFGGPRICAGSELAKLEMAVEFIHHLVLKFNWELAEDDQPFAPFPVDF-----PNG 504
 Db 439 -YVPGAGHRCIGENFAVYQIKTIWSTMLRIYERDLIDG-----YEPVNYTMTIMHTPEN 493
 QY 505 LPIRYSR 511
 Db 494 PVIRYKR 500

Db 494 PVIRYKR 500

Search completed: March 30, 2003, 12:14:58
 Job time : 54 secs

RESULT 15
 US-09-909-903-18
 ; Sequence 18, Application US/09909903
 ; Patent No. US20020052031A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Waterman, Michael R.
 ; TITLE OF INVENTION: Mycobacterium tuberculosis CYP51 POLYPEPTIDES AND
 ; TITLE OF INVENTION: NUCLEIC ACIDS AND THERAPEUTIC AND SCREENING METHODS
 ; FILE REFERENCE: Attorney Docket No. US20020052031A1 1242-17
 ; CURRENT APPLICATION NUMBER: US/09/909,903
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 18
 ; LENGTH: 503
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-909-903-18

Query Match 9.3%, Score 249; DB 10; Length 503;
 Best Local Similarity 24.1%, Pred. No. 2.5e-12;
 Matches 132; Conservative 87; Mismatches 212; Indels 116; Gaps 28;

QY 9 LIPILLPSLST-LLFILLKRRNRKTRFNLPCK-----SGMPFIGETIGYLRPT 60
 Db 26 LLSMLLACAFLLSVLIRLAAGH--LVQLPAGVKSPYFSPFPLGHAIAFGKSPI 82
 QY 61 ATTLDPMQOHVSKYKGYRSNLGCE--PTIVSADAGLNRFILQNEGRLEFCSYPSISIG 118
 Db 83 -----EFLNAYEKYGFPSFTWVGKTFYLLGSDAALLFNSKMDLNAEDVYSLTTP 137
 QY 119 ILGKSMVLVYGD--HRDMSISLN--FLSHARLRTILKDYERHTLVFLVDSMOON- 171
 Db 138 VEGKVAADVDPVPLDQKMLKSGLNIAHFQHVSI-----LEKETKEFEESMGESG 190
 QY 172 --SIFSADDEAKKTFNLMAKHMSMDPEEETDOLKE---YTFMKGVVASAPLNLPG 225
 Db 191 EKNVEBALSE---LIITLASHCLH--GKETRSQLNKRVAGLVADLDGFSHAAMLPLG 243
 QY 226 ---TAVHKALOSRAATILKFERKMEERKIDKEDEEVEVKTDEAEKMSKSDHVRKOR 281
 Db 244 WLPPLPSFRRRDRRAHREIDIFYKAQKRR-----QSOEKI----- 278
 QY 282 TDDLLGNVLKHS-----NLSTEQILDLILSLFAGHETSSVALALAFLOACPRAVEE 336
 Db 279 --DDILQTLTATYKDGRLPTDDEVAGMLIGILLAGHTSSTTSAMGFEL-ARDKTLQ- 334
 QY 337 LREHLEIARAKKELGES--ELNMWDYKKMDFTQCVINETLRIGNVVAFELHRKALKDVRX 394
 Db 335 -KKCYL--OKTVCGENLPLTYDOLKDLNDRICKETLRLPPIIMMRMARTPQTV 390
 QY 395 KGYDIPSGMKVY--PVISAVHLDN--SRYDOPNLFPNPMOONNGASSSGSFSWTGN 450
 Db 391 AGTTPPGHVCSPVNOBKDSWVERLD-----FNPDRILODN---PASGE-KFA----- 438
 QY 451 NYMPFGGPRICAGSELAKLEMAVEFIHHLVLKFNWELAEDDQPFAPFPVDF-----PNG 504
 Db 439 -YVPGAGHRCIGENFAVYQIKTIWSTMLRIYERDLIDG-----YEPVNYTMTIMHTPEN 493
 QY 505 LPIRYSR 511

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on:

March 30, 2003, 12:04:26 ; Search time 49 seconds
(without alignments)
1006.469 Million cell updates/sec

Title:
US-09-502-426a-2

Sequence: 1 METEHTHTLLPLLLPLSLLS.....FAFPVDPNGILIKVSRIL 513

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_73:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2677	99.9	513	2 T46143	steroid 22-alpha-h
2	978.5	36.5	472	1 S55379	cytochrome P450 cy
3	866	32.3	512	2 H96759	probable steroid 2
4	860	32.1	457	2 D85429	cytochrome P450 11
5	815	30.4	382	2 T48613	hypothetical prote
6	761.5	28.4	464	2 T07859	cytochrome P450 ho
7	632	23.6	457	2 T04444	cytochrome P450 -
8	611	22.8	482	2 T02739	probable cytochrom
9	581	21.7	487	1 C71417	cytochrome P450 d1
10	579.5	21.6	485	2 A84859	probable cytochrom
11	579.5	21.5	490	2 H86185	hypothetical prote
12	577	21.5	455	2 T48973	cytochrome P450-11
13	547.5	20.4	489	2 B84733	probable cytochrom
14	497.5	18.6	460	2 D96813	hypothetical prote
15	491.5	18.3	519	1 T02263	cytochrome P450 DW
16	489.5	17.9	444	1 T04602	cytochrome P450 ho
17	479.5	17.3	349	1 S75761	cytochrome P450 -
18	369	13.8	349	2 A86329	Fl4p1.4 protein -
19	357	13.3	518	2 T20908	hypothetical prote
20	338	12.6	453	2 C83722	cytochrome P450 hy
21	329.5	12.3	517	2 T20907	hypothetical prote
22	324.5	12.1	520	2 T24778	hypothetical prote
23	321.5	12.0	518	2 T24783	hypothetical prote
24	316	11.8	500	2 T04737	hypothetical prote
25	310.5	11.6	520	2 T24777	cytochrome P450 ho
26	310	11.6	500	2 T52175	hypothetical prote
27	306.5	11.4	504	2 A25222	cytochrome P450 mo
28	302.5	11.3	491	2 S31277	cytochrome P450 3A
29	300	11.2	491	2 I84735	cytochrome P450 2B
					testosterone 16a-h

ALIGNMENTS

RESULT 1

T46143

steroid 22-alpha-hydroxylase (DWF4) - Arabidopsis thaliana

N:Alternate names: protein T3A5.40

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 16-Feb-2001

C:Accession: T46143

R:Biocker, H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Scheller, C.; Quettler, F.; S

submitted to the Protein Sequence Database, December 1999

A:Reference number: 223024

A:Accession: T46143

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-513 <BLO>

A:Cross-references: EMBL:AL132979

A:Experimental source: cultivar Columbia; BAC clone T3A5

A:Genetics:

A:Map position: 3

A:Insertions: 74/2; 182/3; 233/3; 338/3; 369/3; 396/1; 432/3

A:Note: T3A5.40

C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: heme; iron; metalloprotein

F:308-484/Domain: cytochrome P450 homology <P45>

F:462/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match

Best local similarity 99.9%; Score 2677; DB 2; Length 513;

Matches 512; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	METEHTHTLLPLLLPLSLLSLILFLILKRRNRKTRFNLPGKSGWPLGRTIGLKYPT	60
DB	1	METEHTHTLLPLLLPLSLLSLILFLILKRRNRKTRFNLPGKSGWPLGRTIGLKYPT	60
QY	61	ATTLDGFMQOHVSKYKGYRSLNLFGEPTIVSADAGINFTILONEGRLECYSPISIGIL	120
DB	61	ATTLDGFMQOHVSKYKGYRSLNLFGEPTIVSADAGINFTILONEGRLECYSPISIGIL	120
QY	121	GKMSLVLVGDHNRDMRSISLNFSLSHARLRTILKDVRRHLFVLDSQONSIFSADDEA	180
DB	121	GKMSLVLVGDHNRDMRSISLNFSLSHARLRTILKDVRRHLFVLDSQONSIFSADDEA	180
QY	181	KFTFNLAQKIHMSDPEEETEDLKKRYTFMKGVSAPLNLPGTAVHKAQSRATILK	240
DB	181	KFTFNLAQKIHMSDPEEETEDLKKRYTFMKGVSAPLNLPGTAVHKAQSRATILK	240
QY	241	FIERMEERKLDIKEEDDEEVKTEDEAEKSKSDHVRKQRTDDLLGWLKHSNLSTEQ	300
DB	241	FIERMEERKLDIKEEDDEEVKTEDEAEKSKSDHVRKQRTDDLLGWLKHSNLSTEQ	300
QY	301	ILDLLSLFLAGHETSSVALALAFLLQACPKAVEELREELFETARAKKELGSEELNWD	360
DB	301	ILDLLSLFLAGHETSSVALALAFLLQACPKAVEELREELFETARAKKELGSEELNWD	360

RESULT 4

D85429
cytochrome P450 like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Apr-2001
C:Accession: D85429
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: D85429
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-457 <STO>
A:Cross-references: GB:NC_001268; NID:97270586; PIDN:CA80304.1; GSPDB:GN00140

A:gene: AT4g36380
A:map position: 4
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:396/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 32.1%; Score 860; DB 2; Length 457;
Best Local Similarity 36.8%; Pred. No. 3,3e-48;

Matches 179; Conservative 93; Mismatches 160; Indels 54; Gaps 8;

```

OY 39 LPPGKSGWPLGEGTGYL-----KPYATTTGDPWQOHVSKYKITYNSLFGPEPTVSA 92
DB 2 IINGSLGMYVIGETLNFACGYSSRPVT-----FMDKRKSLGYGVFKNIIGTPIIIST 55
OY 93 DAGLNFFILQNGRLEFECYSPRSIGILGKMSLVLVGDHMDMSISLNFISHARKTI 152
DB 56 DAENVKVVYLVQNGHNFVPRPYSITELIGENSILSINGPHOKRITLIGAFIRSHLDR 115
OY 153 LKQVDRHTLVLDWQONSIFSAODEAKFTFNLMAHIMSMDGEEETOLKEEYTF 212
DB 116 ITRDIEASVYLTASWADLPVHVODEIKMTFELIVYLMSTSPG-EDMNTLKLEPEEF 174
OY 213 MKGVVASPLNLTGTAVHKAQSRATILFERKMERKLDIKEEQQEEVEEKT-----E 266
DB 175 IKGILCIPKPFGRITLYSKAKERLIKMKVVEROVAMTTSPANDVAVILLRDGSD 234
OY 267 DEAEKSKSDHVRKQRTDDLLGWLKHSNLSLSTEDLLILSLFAGHETSSVAIALATF 326
DB 235 SEKQSPDDEVSFGK-----IYEMIPGEETMPTAMTLANKF 270
OY 327 LOACPKAVEELREHLEIARAKKEISELNMWDYKKADFTQCVINETLRGNVRFILR 386
DB 271 LSDNPFVALAKVEEMKRRKRLLEGE-EYKMTDYSLSFTQYINETLRMANIINGVWR 329
OY 387 KALKDVRKGYDIPSGMKVLPVISAHLNDSRYDOPNLFNPRMOQONGSSGSSGFS 446
DB 330 KALKDVEIKGYLIPRGWCVLASFVSHDEDIYNPYQFDPWRDRINGSNSSIC---- 385
OY 447 TWGNMYMFGGPRICAGSELAKLEMAVFIHLVLFKNMELAEDDOPAPFPVDFPGLP 506
DB 386 -----FPPFEGGGRCLCPGLSLKLEISIFLHLVTRYSW-TAEDELYSFPYVMKRRLP 439
OY 507 IRVRSRI 512
DB 440 IRVATV 445

```

RESULT 5

T48613
hypothetical protein F18022.190 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48613
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224493
A:Accession: T48613

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-382 <BEV>

A:Cross-references: EMBL:DB

A:Experimental source: cultivar Columbia; BAC clone F18022

C:Genetics:

A:Map position: 5

A:Introns: 31/2; 139/3; 204/1; 224/3; 253/3; 280/1; 315/3; 349/2

A>Note: F18022.190

Query Match 30.4%; Score 815; DB 2; Length 382;
Best Local Similarity 34.2%; Pred. No. 2.1e-45;

Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

```

OY 45 GMPFEGTIGYKPYTATTTGDMQOHVSKYKITYNSLFGPEPTVSDAGINREILONE 104
DB 2 GMPFEGTISFEKPHRSISIGTFLQORVSRGKVKYSNIGGKAVSCDQELNFIQNE 61
OY 105 GRLFCYSPRSIGILGKMSLVLVGDHMDMSISLNFISHARKTILLADVERHTLFV 164
DB 62 GRLFTSDPKAMHDILGKYSLLATGEIRKIKYIISFINDTSKPDFLCAENLSI 121
OY 165 LSWQONSIFSAODEAKFTFNLMAHIMSMDGEEETOLKEEYTFPMKGVASPLNLP 224
DB 122 LKSWNCRVEVEFKKVKITFLSVWVQNLISIKPEDPARLVYLODFLSMKGFISLPDLP 181
OY 225 GYAVHKAQSRATILFERKMERKLDIKEEQQEEVEEKTDEAEKSKSDHVRKQRTDD 284
DB 182 GTGYNAIKVSNRNRIHONATIEDNNAIREDPLDLSINED----- 224
OY 285 DLGQVNLKHSNLSLSTEDLLILSLFAGHETSSVAIALATFLOACPKAVEELREHLEI 344
DB 225 -----EEHAAI 230
OY 345 ARAKKEISELNMWDYKKADFTQCVINETLRGNVRFILRALKDVRKGYDIPSGWK 404
DB 231 -RANKGDEL-LNMEDQKMEFTQCVISEALRCNINIKYVARKATHDIKEENEYIIPGKW 288
OY 405 VLPVISAHLNDSRYDOPNLFNPRMOQONGSSGSSGSSSTGNNYMPGGPRICAG 464
DB 289 VEPFTAVHNDPSLHNPFEENPRKWKTT-----AFGGGVRCVPG 329
OY 465 SELAKLEMAVFIHLVLFKNMELAEDDOPAPFPVDFPGLP 509
DB 330 GELGKLIAPFLHILVLSYKWKIKSDMPIAHNPVEYKRGMLLEI 374

```

RESULT 6

T07859
cytochrome P450 homolog - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 16-Feb-2001
C:Accession: T07859
R:Bishop, G.J.; Harrison, K.; Jones, J.D.
Plant Cell 8, 959-969, 1996
A:Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes th
A:Reference number: Z16181; MUID:96266705; PMID:8672892
A:Accession: T07859
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-464 <BIS>
A:Cross-references: EMBL:U54770; NID:g1421740; PIDN:AA817070.1; PID:g1421741
A:Experimental source: strain GCR758
C:Genetics:
A:gene: dwarf
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
F:273-436/Domain: cytochrome P450 homology <P45>

Query Match 28.4%; Score 761.5; DB 2; Length 464;
Best Local Similarity 33.4%; Pred. No. 7.9e-42;
Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

OY 12 LILLPFLSLULLFLILKRRN--RKTDFMLPFGKSGWPLGEGTIGYKPYTATTTG-DPM 68

DB 5 LFLSSFFGLCTIFCTALLMNOVKYKQKMLPPTGMPLFGETTEFLK-----LDPSTM 58
OY 69 OOHVSKYKTYISNLFGEPTIVSADAGLNRFILONBGRLEFEGSYPRISGIGLKMSMLVL 128
DB 59 KNOBARVSGFESHILGCTIYSMOSELNRYILYNBAKGLVGYQPSMIDILGKCAIAV 118
OY 129 VGDMDHMSISLNFSLHARLITLLKDYERHTLFVLDWMQONSIFSADEAKKFTFNIM 188
DB 119 NOSAHYKMGALLSTLSPIMRDLPLPKIDEPMSHILTNM-DNKVIDIEKTKMAFLSS 177
OY 189 AHHMSMDGEET---EOLKKEVYTFMKGVASAPLNLGTAVHKLQSRATILKIEK 245
DB 178 LKQI---AGIESTSLAQEFNSFFVLVGTLSLPLNLNTNTHGFOAKKIIIVLRL 233
OY 246 MEERKIDKEEDQEEBEVTEDEAEMSKSDHVRKQRTDLDLGLWVYKHS---NLSTEOI 301
DB 234 IEERR-----ASKEIQHMDGLYLNNEATRKLDDDM 266
OY 302 LDLLISLFAHETSSVAIALAIFLQACPKAVEELREHLEIARAKKEGSELNMDWY 361
DB 267 IDLLITLLSGEYVSTTSMMAVKYLDHPRKYLEELRKHMAIREKKP--EDPIDYNDY 324
OY 362 KKMDFQCYINETLRGLGVVRFHRAKLDVRYKGYDIPSGKVLVVISAVHLDNSRYQ 421
DB 325 RSKRFTRAVILETSRLATVNGVLRKTTODMEINGIITIKGRITVYTRELNIDPRLPD 384
OY 422 PULFNFWMOQONGASSGSGSFTWGNMYPFGGPRLCAGSELAKLEMAVFIHHLVL 481
DB 385 PVSFNPRWMDKS-----LEHNSFLVFGGCTQCGKELGVALEISPLHYFT 433
OY 482 KFNWELAEEDQPFAPFVDFPNCPLRVS 510
DB 434 KYRWEIRGGDKLMKPRVEAPNGLRIRVS 462

RESULT 7

T04444
cytochrome P450 - Arabidopsis thaliana
N:Alternate names: protein T18B16.200; protein T5K18.10
M:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2000
C:Accession: T04444; T05806
R:Byan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.
submitted to the Protein Sequence Database, April 1998
A:Reference number: 215359
A:Accession: T04444
A:Molecule type: DNA
A:Residues: 1-457 <BEV>
A:Cross-references: EMBL:AL021687
R:Experimental source: cultivar Columbia; BAC clone T18B16
R:Byan, M.; Van Der Schuren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, April 1998
A:Reference number: 215453
A:Accession: T05806
A:Molecule type: DNA
A:Residues: 131-457 <BEW>
A:Cross-references: EMBL:AL022580
R:Experimental source: cultivar Columbia; BAC clone T5K18
C:Genetics:
A:Map position: 4
A:Introns: 67/2; 173/3; 302/3; 358/1; 393/3
A:Note: T18B16.200; T5K18.10
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase
F:272-433/Domain: cytochrome P450 homology <P45>

Query Match 23.6%; Score 632; DB 2; Length 457;
Best Local Similarity 31.8%; Pred. NO. 1.8e-33;
Matches 151; Conservative 92; Mismatches 176; Indels 56; Gaps 12;
OY 13 LLLPSLLFLILLKRNKTRNLPFGKSGMPLGTTIGYLAFTATTGADPMQOHV 72

DB 10 LFRAGSLFLYLRCLISQRRFGSSKLPPLPGTMGPVVGST---FOYLSDDP-NVFEQSKQ 65
OY 73 SKYGIYRNSLFGEPITVADAGLNRFILONBGRLEFEGSYPRISGIGLKMSMLVLVGM 132
DB 66 KRGSVFKTHVAGCPVMSLSPRAKFLVLTSHLFKFPFPAKEMMLKQALFFHQGY 125
OY 133 HDMSRISLNFSLHARLITLLKDYERHTLFVLDWMQONSIFSADEAKKFTFNIMAKHI 192
DB 126 HAKRLVLRAPMPEISIRN-NVWDIESIADQSLRWS-EGTMTVWQEMKTYFNVALLSI 183
OY 193 MSMDGEET---EOLKKEVYTFMKGVASAPLNLGTAVHKLQSRATILKIEK 249
DB 184 F---GKDVLYREDIKRYLLEKGYNSMPVLPPTLHKSKAKKELSOILARLSR 239
OY 250 KLDIKEDQEEBEVTEDEAEMSKSDHVRKQRTDLDLGLWVYKHS---NLSTEOI 308
DB 240 R-----QNGSSH-----NDLGSFMDKRELDEQIADNITIGY 272
OY 309 LFAHETSSVAIALAIFLQACPKAVEELREHLEIARAKKEGSELNMDWYKMDFTQ 368
DB 273 IFARPTVASVMILKYLAEINPVLEAVTEBQMAI-RKDEGES-LTWGDTKKMPLRS 330
OY 369 CYINETLRGVVRFHRAKLDVRYKGYDIPSGKVLVVISAVHLDNSRYQPNLFNW 428
DB 331 RVIQETLRVASILSTFRFAVEDVEYGLIPGKWLFLFRNIHSADIFSNPKFDS 390
OY 429 RMQOONGASSGSGSFTWGNMYPFGGPRLCAGSELAKLEMAVFIHHLVL 483
DB 391 RFE-----VAKRPTMFPFGNGTSHCPNELAKLEMSIMIHLLTKY 432

RESULT 8

T02739
probable cytochrome P450 At2g29090 [imported] - Arabidopsis thaliana
N:Alternate names: cytochrome P450 homolog T914.17
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02739; D84692
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.
A:Reference number: 214710
A:Accession: T02739
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-482 <ROD>
A:Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461849
R:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vent
Nature 402, 761-766, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84692
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: GB:AE002093; NID:g3461849; PID:NAC3325.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29090; T914.17
A:Map position: 2
A:Introns: 80/2; 187/3; 238/3; 321/3; 351/3; 413/3; 453/2
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
F:291-453/Domain: cytochrome P450 homology <P45>

Query Match 22.8%; Score 611; DB 2; Length 482;
Best Local Similarity 29.8%; Pred. NO. 4.5e-32;
Matches 153; Conservative 101; Mismatches 192; Indels 68; Gaps 13;
OY 16 PSLLSLFLILLK-----RNRKTRNLPFGKSGMPLGTTIGYLAFTATTGADPMQ 69

Db 20 PALTLITVVVVVLLFKMWMKQRLRLPPSGMLPIGFT---LALTYENP-NSFEA 75
 QY 70 QHVSXGKIYNSNLFGEPTIYSADAGLNRLFLQNEGRLEFESYPSRIGIIGLKWMGLVIV 129
 Db 76 TRQNKRGDIFTHILGCGVMSISPEARMLVSKAHLFKPLPPSKRRMIGPEALFPHQ 135
 QY 130 GDMHEDM-RSTSLNLFNARLRLTLKQVREHTLFLDSMOONSIFSAODEAKKTFFMLM 188
 Db 136 GPHSTLKLRIVSSSPMSALRPTV--SHIELLVLTQTLSSWTSQNSINFLYMKRRAFPDVA 193
 QY 189 AKHMSMDPGEEE---TEOLKREYTFMKGVASAPLNPJTAHYKALOSRAATILKFIK 245
 Db 194 ---IMSAGDKKEEPTIYVILKLYQRLRGYNMPLDPLGLFKHSMARLEISELKV 250
 QY 246 MEERLIDKEEDQEEBEVKTDEAMSKSDHVRKQRTDILLGWYL---KHSNLTSEQ 300
 Db 251 IERRENGREE-----GGLIYGLVLAQDKRNLSDSQ 283
 QY 301 ILDLISLFLAGHETSSVALATFELQACPVAEELEHELEI-ARAKKEGESELNWD 359
 Db 284 IADNLIGVIFATDTTASVLTWMLKYLHDHNLQEVSRQPSIRQIKKE--NRRLSWE 341
 QY 360 DYKMDPTQCVINETLRGLNVRFLHRKALDVRKGYDIPSGWVLPVISAHLNDRY 419
 Db 342 DTRKMPLTTRVIOETLRASVLSFTFREAVQDEVDGLIPKGMKVLPLFRRIHSSSEF 401
 QY 420 DQPLNFNWRQOONNGASSGSGSFSTWGNMNPFGGPRICAGSELAKLEMAVFIHNL 479
 Db 402 PDPERFDSRE-----VAPKPYTMPFGNCHVSCPSSELAKLEMLLHL 448
 QY 480 VLKFNWELAEDDOPFAF-PFVDFPNGLPPIRYSRI 512
 Db 449 TTSFRWEVIGDEGIQYGFPPVKKGLPIRYPTI 482

RESULT 9

C71417
 cytochrome P450 d13695c - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 A:Variety: columbia
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: C71417
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dink
 P.; Medler, H.; Wedder, E.; Mambutt, R.; Weitzmeger, T.; Pohl, T.M.; Terry, N.; Gled
 avanagh, T.; Hempel, S.; Kotter, P.; Entlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
 erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reichen, S.; Ans
 C.; Chalmatzis, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
 A:Reference number: A71400; MUID:98121113; PMID:9461215
 A:Accession: C71417
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-487 <BEV>
 A:Cross-references: GB:Z97338; NID:g2244870; PIDN:CAB10309.1; PID:g2244888
 C:Genetics:
 A:Gene: d13695c
 A:Map position: 4C0D9-4G3845
 C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
 F:433/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match

Best Local Similarity 21.7%; Score 581; DB 1; Length 487;
 Matches 143; Conservative 100; Mismatches 205; Indels 76; Gaps 11;

QY 17 SLTSLILFLILKRRNRKTRFNLPPGSGWPLFETIGYKPYATTATIGPMOAHSKYK 76
 Db 3 SLPLVITFHHVYQWRNPKTKGRLPGSGMGPFIFGETFEFFKPRDALQSTFIKDRVLRFP 62
 QY 77 -----KIYRSNLFGEPTIYSADAGLNRLFLQNEGRLEFESYPSRIGI-----LG 121

Db 63 ADESSILHSFRTSLFGDKALISMDMEIN-----LMAKANSVPQYTKSVIRLFG 112
 QY 122 KWSMVLVGDHNRDMRSLSLFLSHARLRTLLDQVREHTLFLYDSMOONSIFSAODEAK 181
 Db 113 ENNLFLOSKEHSHKVRNLTFFQLDPQGLKSMIDVDLARTVMEGARNGYLDKERTSS 172
 QY 182 KFTFNIAKHIM-SMDPGEETEOUKREYTFMKGVASAPLNPJTAHYKAL----- 232
 Db 173 KILGGLAKKVMGEMEP--EAKELALCWRFGSGWFRFFLNLPGYVYKMKKYLFPQYT 230
 QY 233 -----OSRAATILKFIERKMEERLIDKEEDQEEBEVKTDEAMSKSDHVRKQRTD 286
 Db 231 EADISQARKKMMKLRTVLTFRASGELEFFNIIJEGMEGE----- 274
 QY 287 LGWILKHSNLTSEQDILLISLFLAGHETSSVALATFELQACPVAEELEHELEIAR 346
 Db 275 -----GETSMVENAVEITYTFELVANETPRILAAVVKFISDHPKVKOELOREHEEIVR 328
 QY 347 AKKEIGESELNWDYKMDPTQCVINETLRGLNVRFLHRKALDVRKGYDIPSGWVLP 406
 Db 329 GKAE-KGGGLWEDYKGMHFTQWYNESLRITSTAPVLYKLEHDFQVGDITTPAGWTFM 387
 QY 407 PVISAHLNDRYDQPLNFNWRQOONNGASSGSGSFSTWGNMNPFGGPRICAGSE 466
 Db 388 G-YPHIHFNSEKYEDPAFNFWRWEGKDLGAIYS-----KTFIFGAGRILCYGAE 437
 QY 467 LAKLEMAVFIHNLKFNWELAEDDOPFAF-PFVDFPNGLPPIRYSRI 510
 Db 438 FAKQMAVFIHNL-FRYRMSKSGTITIRSFMLPFGGCDVOIS 480

RESULT 10

A84859
 probable cytochrome P450 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 23-Mar-2001
 C:Accession: A84859
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bento, M.I.; Tom, C.D.; Fujii, C.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
 euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: A84859
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-485 <STO>
 A:Cross-references: GB:AEO02093; NID:g4512670; PIDN:AAD21724.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g42850
 A:Map position: 2
 C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: heme; iron; metalloprotein
 F:432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match

Best Local Similarity 21.6%; Score 579.5; DB 2; Length 485;
 Matches 147; Conservative 107; Mismatches 203; Indels 53; Gaps 11;

QY 12 LLLPLSLSLFL-----ILKRRNRKTRFNLPPGSGWPLFETIGYKPYAT-TL 64
 Db 14 LCIATWISLTLFFRRKKNHRTYKKIQKKK--LLPGEMGLPIGEMDPFYKQNSNRVF 71
 QY 65 GDFMQHVSXGKIYNSNLFGEPTIYSADAGLNRLFLQNEGRLEFESYPSRIGIIGKMS 124
 Db 72 EDVFNRIIRKHGIFETRLMGSPFTIYNGANRLTLSEFSLVSSWSSSSVOLGMNC 131
 QY 125 MLVLVGDHNRDMRSLSLFLSHARLRTLLDQVREHTLFLYDSMOONSIFSAODEAKKT 184
 Db 132 IMAKQEKHRVLRGIANSLSYIGLESLIPKLDQYKFNHEWRKKEISILYRSKAVLT 191
 QY 185 FNMAKHI--MSMDPGEETEOUKREYTFMKGVASAPLNPJTAHYKALOSRAATILKEI 242

Db 192 FVVEECIIGIVEIGMLEV-----FERNLEGVALPVEFPCSFARKKKRLIEITFL 245
243 ERKMEERKIDKEEDOEVEEVKTEDEAENS SDHVRKORTDDLLGWLKHSNLSSTEQITL 302
246 VGVKREKREMEKEGAE-----KPNITLFSVLVELLKGV-----ITEEVV 287
303 DILSLFPGHETSVAATLALFPIQACPKAVEELREHELEIARAKKEGSE-LMWDY 361
288 DMVVLVFAHDTSTYAMSMTEKMLAQHPTCRDTLLOEHAQT---KANKGEYLVLEDV 344
362 KKMDFTCQYINETLGNVRFELHRRKALVDYKGYDIPSGKVLVPIASVLDNSRYQD 421
345 KMKXSMQVYRRTMLSPFIIEGSRKAVADIDYGETTIPKMKILMTYTGTHYNEPIOD 404
422 PVLNPMWQOQNGASSSGSFGSTWNNYMPFGGPRLCAGSELAKLEMAVFIHHLVL 481
405 PMSFPTREKRIQAVT-----YLPFGGPRLCAGHQIAKISILVFMHVVYT 451
482 KFMWELAEODOPFAFPVDFPN-GLPIRVS 510
452 GFDWSLYPDEFITSMDFPPLGMPFIKIS 481

RESULT 11

H86185
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86185
R:Theologos, A.; Becker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nucleotide 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matil, R.; Matzfall,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shino, P.; Southwick, A.M.; Sun, H.; Tjallou,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H86185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <STO>
A:Cross-references: GB:AE005172; NID:g2388581; PIDN:AA071462.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 21.6%; Score 579.5; DB 2; Length 490;
Best Local Similarity 28.6%; Pred. No. 5e-30;
Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;
34 KTRFNLPGKSGMPFLEGTIGYLPYATYATLGDPMOQHVSKYK--IYRSNLFGEPTIYS 91
40 ENRHLLPREDGLGMPITIGMNLFLRAFKTSDDPSFTRIKRGPRGTYKAAHFGNPSTIV 99
92 ADAGNRLFLONEGRLEFECSPYRSIGILGKMSLVLDGMDHRMRSISLNFSLHARLRT 151
100 TSDPCRVLITDDD-AFKPGMPTSMELIGRKSFGVIGSEEHKRLRLTLAIVNGHEALS 158
132 ILLKVERHTLFLVDLSWQONSIFSNODEAKKFTFNMAKHIMSMDPGEETEQLKKEYT 211
159 TYPIPIEENVTIVDKMKMGFEFLTLHLRKLTFRII--MYFLSESESEVMDALEREVTA 217
212 FMKGVVAPLNLPGTAYKALQSRAITLKTIRKMEERKLDIKEDOEVEEVKTEDEAEM 271
218 LNYGVRAAVNLPGFAIHALAKRTLVAAFGSIVTER-----NQRKNTLSKKDKML 271
272 SKSDVRKORTDDLLGWLKHSNLSSTEQITLILSLFAGHETSVAATLALFPIQACP 331
272 DNLNLVVK-----DEK-----GKTLDEDEIIDLVLMTYLNAGHSGHTIMATVTLQENR 320

QY 332 KAVEELREHELEIARAKKEGSELNWDYKKMDFTCQYINETLRLGNVRFELHRRKALD 391
Db 321 EYLQRAKAEQOEIMILKSREPGQKGLSLKTRKMEFLISQVDETLRTVTFSLAFRAKTD 379
QY 392 VRYKGYDIPSGKVLVPIASVLDNSRYQDPLFNPMWQOQNGASSSGSFGSTWNNY 451
Db 380 VMANGTILIPKGVVLVWFRVDVIDPEVFPDPKRPDAR--DNG-----FVPRKA 428
QY 452 YMPFGGPRLCAGSELAKLEMAVFIHHLVLKFMWELAEODOPFAF 496
Db 429 FLFPAGSHLCGNDLAKLEISILFHLFLTKYQVRSNECEVMY 473

RESULT 12

T48973
cytochrome P450-like protein - Arabidopsis thaliana
N:Alternate names: protein P14D17.40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T48973
R:Jordan, N.; Banger, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd,
submitted to the Protein Sequence Database, April 2000
A:Reference number: 225008
A:Accession: T48973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <JOR>
A:Cross-references: EMBL:AL353992; GSPDB:GN00061; ATSP:F14D17.40
C:Genetics:
A:Experimental source: cultivar Columbia; BAC clone F14D17
A:Gene: ATSP:F14D17.40
A:Map position: 3
A:Introns: 70/2; 178/3; 229/3; 312/3; 343/3; 346/1; 381/3
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 21.5%; Score 577; DB 2; Length 455;
Best Local Similarity 28.9%; Pred. No. 6.6e-30;
Matches 145; Conservative 84; Mismatches 173; Indels 100; Gaps 10;
32 NKTRFNLPGKSGMPFLEGTIGYLPYATYATLGDPMOQHVSKYKRYNSLFGPTIYS 91
28 NPKSNGKLPRGSMGPIIGETIDPFKPYGFYISLYLKMKRLRYGRLPFTNLGKYTVS 87
92 ADAGNRLFLONEGRLEFECSPYRSIGILGKMSLVLDGMDHRMRSISLNFSLHARLRT 151
88 TDKVNMELIROENKSFILSYDGLMKRPGKDSLFLKIGNIKHKKIQTILHLSSEGLR 147
152 ILLKVERHTLFLVDLSWQONSIFSNODEAKKFTFNMAKHIMSMDPGEETEQLKKEYT 211
148 KILKMDRYTRRHLSKAKTGRLDVKDAVSKLIILHLPKMS-----NLKPQTQA 198
212 FMKGV-----VSAPLNLPGTAYKALQSRAITLKTIRKMEERKLDIKE 255
199 KLMGIFKAFTEPFRFTSYLISAGKLYNTLW--ACREGKREIKDITYAKTSEBK---- 252
226 EDQEEVEKTEDEAEMSDHVRKORTDDLLGWLKHSN-----LSTEQITLILSLF 310
253 -----DDFLNTAIEESEKAGELINERVAITTLITLISC 284
311 AGHETSVAATLALFPIQACPKAVEELREHELEIARAKKEGSELNWDYK-KMDFTCQ 369
285 VTQDTSTAIICLAVFLLENPRVLAELEKHE--EVLLSESEDEGEGVTEWEEYHKKTFTN 342
370 VINETLRLGNVRFELHRRKALVDYKGYDIPSGMVLVPIASVLDNSRYQDPLFNPM 429
343 -----MKSGTIPRAGWLVMIIPSVVHNDRELTEMPFEFNPR 379
430 WQOQNGASSSGSFGSTWNNYMPFGGPRLCAGSELAKLEMAVFIHHLVLKFMWELAE 489
380 WEGKELRAGS-----KTFVFGTGLRQCAGAEFARLQISVLEHNLVLTYNFSLHQ 429
490 DDOPFAFPVDFPNGLPIRVS 511

DB 430 DCEVLRVPAHLPNGISINISK 451

RESULT 13

probable cytochrome P450 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: B84733

Rilin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; enus, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84733

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1489 <STO>

A:Cross-References: GB:AE002093; NID:g3831452; PIDN:AAC69934.1; GSPDB:GN00139

C:Genetics:

A:Gene: Atg32440

A:Map position: 2

C:Superfamily: synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 20.4%; Score 547.5; DB 2; Length 489; Best local Similarity 27.6%; Pred. No. 5,9e-28; Matches 146; Conservative 105; Mismatches 211; Indels 67; Gaps 14;

QY 8 TLPLLLPSLLFLI-LIKRRN-----RKRPNLPKSGMPLGTTIGYLK 57
 DB 4 TGLILMWF-LILGLFVLKWLKRVNWIYVSKLGEKKHYLPDGLGWMVIGNMWSFLR 62
 QY 58 PYATATLGDPMOQHSYK--TYRSNLGEPITVSADAGLNRFITLONEGRLEPCSVPR 115
 DB 63 AFTSDPESTQSTYIRIGTGTYIKAHMEYPCVLTPTPTCRVLTDDO-AHHIGPKS 121
 QY 116 IGGILKWSMLVYGDHMDRMSISLNFSLHARLITLLKDVHRTFLVDSMOQNSIFS 175
 DB 122 TMLIGKSSVGSIFPEHKKRLRLTSAVNGPEALSYIGFIEVTVDLEKMSKMEIE 181
 QY 176 ADEAKKFTTNMAKIHMSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAHKAQSR 235
 DB 182 FLSHLRKLTPEKVI-MYTFISSESEHVDLSERETNLNYSVGRAMGILPGFAVHRAKAR 240
 QY 236 ----ATLKRERKMERKIDI--KEEDQEEVEKTEDEAMSKSDHVRKQRTDDLLGW 289
 DB 241 KTLVAFOSTIVTRNRNOKONISSNRKMDLNDLIDVDE-----NGRVLD-- 286
 QY 290 VLKHSNLSTEQIIDLILSLFAGHETSSVAIALAIFLOACPKAVEELREHLEIARAK 349
 DB 287 -----BEIIDLILMYINAGHESGHLTMATILMOHPHILQAKKEQBEI--YKK 335
 QY 350 ELGSESLNMDYKRMDFQTCVINETLRLGNVRLHRAKLDVRYKGYDIPSGKVLPIV 409
 DB 336 RAGQOKLTLEKREMYLTSQVIDETTLVITFSLTAFREAKSDVQMDYIIPKGAIVTW 395
 QY 410 SAYHLNLSRYDQPLFMWPMOQONNGASSGSGSFSTWNNVWPEGGRPLGSELA 469
 DB 396 RNVLHDEITPDPKKPEPSRME-----GYTPKAGTFLPGLSGHLLCGNDLAK 443
 QY 470 LEMAVFTIHLVLKFNWELAEEDQFAFPVDFPGLP-----IRVSRIL 513
 DB 444 LEISIFLHFLKRVKRSNPGCVMF-----LPHNRKDKCLAIITRTM 488

RESULT 14

hypothetical protein T30F21.17 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 19-Apr-2002

C:Accession: D96813

R.Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Hulzar, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakao, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D96813

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1460 <STO>

A:Cross-References: GB:AE005173; NID:g4836883; PIDN:AAD30586.1; GSPDB:GN00141

C:Genetics:

A:Gene: T30F21.17

A:Map position: 1

C:Superfamily: synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 18.6%; Score 497.5; DB 2; Length 460; Best local Similarity 26.2%; Pred. No. 9,3e-25; Matches 133; Conservative 102; Mismatches 205; Indels 67; Gaps 12;

QY 13 LLPLSLILLFLILKRRNRKTRFNLPPKSGMPLGTTIGYLKPYATATLGDPMOQHV 72
 DB 9 MLWALVAVNLSHWLWNSNPKCGKLPDSMGPIGTEITLDFPKGCVGIPFVVKRM 68
 QY 73 SKYGIKTRSNLGEPTIVSADAGLNRFITLONEGRLEPCSVPRISGGLKWSMLVYGD 132
 DB 69 IRRGPLERTNIFISKVYSDPDVHQIFQENTSEFLGPDYFVKYFGDNLEKEVFI 128
 QY 133 HRMRSISLNFSLHARLITLLKDVHRTFLVDSMOQNSIFSADAKKFTTNMAKHI 192
 DB 129 HKYLOKTRMILGSEGAKOMLGMMDKATDHRISASGSPFNKREVENLVAAYMPKL 188
 QY 193 MSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAHKAQSRATILKFERKMERKLD 252
 DB 189 ISNLKPEQSKLIDN-----LNANFNDWPKSF-LRSTWKAVTKALKSR-- 231
 QY 253 IKEDQEEVEKTEDEAMSKSDHVRKQRTDDLLGVLKH-----SNLSTEQIIDLILS 307
 DB 232 -----EALIQWKVDLMKRRETKRQ--EDFLNTLLELEKDGSGFPDGSAINILFL 281
 QY 308 LIFAGHETSSVAIALAIFLOACPKAVEELREHLEIARAKKESESLNMDYK-KMDF 366
 DB 282 LAPLRGCTSSCTALAVKFSKDPKVLAEKREKALVYDNKRD-KEAGVSNEEYRHNMTF 340
 QY 367 TOCVINETLRLGNVRLHRAKLDVRYKGYDIPSGKVLPIVSAVHLNLSRYDQPLF 425
 DB 341 TNWVSNVLRANTPTPLFRKAVQDVEIKKY-----YLEVIMHGRG----- 382
 QY 426 NPMWQOONNGASSGSGSSTWGN-NYMPGGGPRCAGSELAELKEMAVFTIHLVLKFN 484
 DB 383 --WLMQ-----GKEMWGSKTFMAFGYVRLCGAESRLQMAIFLHLVAYD 429
 QY 485 WELAEEDQFAFPVDFPGLPPIRVSR 511
 DB 430 FSWQDSEIIRSPHQTOKLLINISQ 456

RESULT 15

cytochrome P450 DMARP3 - maize

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Zea mays (maize)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: T02263

R:Winkler, R.G.; Helentjaris, T.

Plant Cell 7, 1307-1317, 1995

A>Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibbe

A:Reference number: Z14648; MUID:96004534; PMID:7549466

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2003, 09:02:00 ; Search time 38 Seconds

(without alignments)
559,930 Million cell updates/sec

Title: US-09-502-426a-2

Perfect score: 2681
Sequence: 1 MFETEHHTLPLLLPSLLS.....PAPPVDFNGDPIRVSRIL 513

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	978.5	36.5	472	C901_ARATH	Q42569 arabidopsis
2	864	32.2	524	C90C_ARATH	G9m066 arabidopsis
3	761.5	28.4	464	C985_LYCES	Q43147 lycopersico
4	579.5	21.6	480	C983_ARATH	Q23051 arabidopsis
5	503	18.8	492	CP26_BRARE	P79739 brachydanio
6	491.5	18.3	519	C981_MAIZE	Q33246 zea mays (m
7	479.5	17.9	444	C120_SYNY3	Q59950 synecocyst
8	461	17.2	497	CP26_MOUSE	O55127 mus musculu
9	446	16.6	512	C26A_HUMAN	O9nrf3 homo sapien
10	444.5	16.6	492	CP26_XENLA	O93323 xenopus lae
11	439.5	16.4	492	CP26_CHICK	O9pud4 gallus gall
12	438	16.3	497	CP26_HUMAN	Q43174 homo sapien
13	324.5	12.1	520	YRV5_CAEEL	Q27514 caenorhabdi
14	321.5	12.0	518	YRVA_CAEEL	Q27510 caenorhabdi
15	316	11.8	500	C912_ARATH	O65780 arabidopsis
16	312.5	11.7	504	CP33_RAT	P05183 rattus norv
17	312	11.6	501	CPJ5_MOUSE	O54749 mus musculu
18	310.5	11.6	520	YRV1_CAEEL	Q27513 caenorhabdi
19	303	11.3	501	CPJ6_MOUSE	O54750 mus musculu
20	302	11.3	494	CPA8_MESAU	P24454 mesocricetu
21	299.5	11.2	492	CPB8_MOUSE	P33272 rattus norv
22	297	11.1	491	CPB8_MOUSE	P12760 mus musculu
23	294.5	11.0	491	CPBA_RABIT	P00178 oryctolagus
24	294	11.0	491	CPB1_RAT	P00176 rattus norv
25	293	10.9	491	CPB1_RAT	P10614 saccharomyc
26	291.5	10.8	530	CP51_YEAST	P48421 arabidopsis
27	290.5	10.8	502	C83A_ARATH	P51588 rattus norv
28	289	10.8	503	CP39_RAT	O55071 mus musculu
29	288.5	10.8	492	CPB3_MOUSE	P05184 homo sapien
30	288.5	10.8	503	CP33_HUMAN	O95078 drosophila
31	288.5	10.8	459	CP18_DROME	P29980 anabena sp
32	288	10.7	538	CPXN_ANASP	Q27517 caenorhabdi
33	287.5	10.7	520	YRV5_CAEEL	

34	286.5	10.7	533	1	CP51_CANGA	P50859 candida gla
35	285.5	10.6	502	1	CPJ2_HUMAN	P51589 homo sapien
36	284.5	10.6	505	1	C762_SOAME	P37122 solanum mel
37	282	10.5	514	1	C941_VICSA	O81117 vicia sativ
38	281.5	10.5	499	1	C831_ARATH	O65782 arabidopsis
39	281.5	10.5	504	1	CP3G_MOUSE	O64481 mus musculu
40	281	10.5	528	1	CP51_CANTR	P14263 candida tro
41	280.5	10.5	504	1	CP3B_MOUSE	O64459 mus musculu
42	280	10.4	499	1	C771_SOAME	P37123 solanum mel
43	280	10.4	500	1	CPJ1_RABIT	P52786 oryctolagus
44	279.5	10.4	513	1	C773_SORBN	O48928 glycine max
45	279.5	10.4	519	1	YRV8_CAEEL	Q27520 caenorhabdi

ALIGNMENTS

```

RESULT 1
ID C901_ARATH STANDARD: PRT: 472 AA.
AC Q42569;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 90A1 (EC 1.14.-.-).
GN CYP90A1 OR CYP90 OR CPD OR A5G05690 OR MJ3.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC eustosids II: Brassicales: Brassicaceae: Arabidopsis.
OC NCBI-TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=96200769; PubMed=8612270;
RA Szekeres M., Nemeth K., Koncz-Kalman Z., Mathur J., Kauschmann A.,
RA Altmann T., Redei G.P., Nagy F., Schell J., Koncz C.;
RT "Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
RT controlling cell elongation and de-etiolation in Arabidopsis.";
RL Cell 85:171-182(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT clones.";
RL DNA Res. 4:215-230(1997).
RN [3]
RP -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: X87367; CA60793.1; -
DR EMBL: X87368; CA60794.1; -
DR EMBL: AB005237; BAB09663.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PROSITE: PS00066; CYTOCHROME_P450; 1.
KW Oxydoreductase, Monooxygenase, Transmembrane, Heme, Multigene family.
FT TRANSMEM 1
FT BINDING 418 418 HEME (BY SIMILARITY).
FT SQUENCE 472 AA; 53785 MW; 41A73F46D4E343F CRC64;
Query Match 36.5%; Score 978.5; DB 1; Length 472;
Best Local Similarity 41.9%; Pred. No. 1.7e-51;

```

Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLLSLSLSLLFLILKRRNKRTRNPPGSGWFLGCTTGYLKPATYATLGDGMOOH 71
 7 LLLLSLSLSLSLLFLILKRRNKRTRNPPGSGWFLGCTTGYLKPATYATLGDGMOOH 71
 Db 7 LLLLSLSLSLSLLFLILKRRNKRTRNPPGSGWFLGCTTGYLKPATYATLGDGMOOH 71
 QY 72 VSKYGVYNSNLFGEPTTYSADAGLNRTFLONRGRLFECSYRSIGTILGKMSMLVYGD 131
 64 VARKSVFTNHLFGEPTTYSADAGLNRTFLONRGRLFECSYRSIGTILGKMSMLVYGD 131
 Db 64 VARKSVFTNHLFGEPTTYSADAGLNRTFLONRGRLFECSYRSIGTILGKMSMLVYGD 131
 QY 132 MHRDMRSLSNLFNARLRTILKDRERTFLVDLSMOONSIFSADDEAKKFTFNLMANH 191
 124 LKRRHSLTMSFANSIIKDHMLDRLVRFNLDMSWSRVLV - MEAKKFTTFLTYVQ 181
 Db 124 LKRRHSLTMSFANSIIKDHMLDRLVRFNLDMSWSRVLV - MEAKKFTTFLTYVQ 181
 QY 192 IMSNDPGEETQLKKEVTYFPMKGVASAPLNPGTAHKALOSRTATLTKERKEERL 251
 182 LMSFDPG - EWSLSRKEYLVLEEGFSLPLFSTTYRKAIDAR - - - - - RKVAEALT 232
 Db 182 LMSFDPG - EWSLSRKEYLVLEEGFSLPLFSTTYRKAIDAR - - - - - RKVAEALT 232
 QY 252 DIKEDEQEEVKTEDEAMSKSDHVRKQRTDDLLGWLKSNLSSTEQILDLISLFLA 311
 233 VVVKRRREEEGAE - - - - - RKKOMLALALADDFSDSEITVFLVALVA 278
 Db 233 VVVKRRREEEGAE - - - - - RKKOMLALALADDFSDSEITVFLVALVA 278
 QY 312 GHETSSVALAIFLOACPKAVEELREHLEIARAKKEGSELMNDYKKMDFTQCVI 371
 279 GYETSTIMTLAVKFLTERPLALQKEHEKIRAMKSD - - - - - SYLSMDSYKSMPTQCV 336
 Db 279 GYETSTIMTLAVKFLTERPLALQKEHEKIRAMKSD - - - - - SYLSMDSYKSMPTQCV 336
 QY 372 NETLRLGNVRLKRLKADVKRYKGDIDSGKVLVSAVHLDSRYDQPLFNMWMO 431
 337 NETLRLGNVRLKRLKADVKRYKGDIDSGKVLVSAVHLDSRYDQPLFNMWMO 431
 Db 337 NETLRLGNVRLKRLKADVKRYKGDIDSGKVLVSAVHLDSRYDQPLFNMWMO 431
 QY 432 OONNGSSSGSGSFSGMGN - YMPFGGPRLCAGSELALMAVFNHLYLKFNELAD 490
 397 - - - - - SNSVTGSSNVTTPFGGPRLCAGSELALMAVFNHLYLKFNELAD 446
 Db 397 - - - - - SNSVTGSSNVTTPFGGPRLCAGSELALMAVFNHLYLKFNELAD 446
 QY 491 DQFAEPFVDFPNGLPTRVSR 511
 447 DLVFEPTTRTQKRYPIFYKR 467
 Db 447 DLVFEPTTRTQKRYPIFYKR 467

RESULT 2

C90C_ARATH STANDARD: PRT; 524 AA.

AC Q9M066; 023242;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome P450 90C1 (EC 1.14.-.-) (ROTUNDIFOLIA3).
 GN ROT3 OR CYP90C1 OR AT4G36380 OR C7A10.980 OR F23E13.220.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosidae;
 CC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
 CC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98361880; Pubmed=9694802;
 RA Kim G.-T., Tsukaya H., Uchimiya H.;
 RA "The ROTUNDIFOLIA3 gene of Arabidopsis thaliana encodes a new member
 RT of the cytochrome P-450 family that is required for the regulated
 RT polar elongation of leaf cells.";
 RL Genes Dev. 12:2381-2391(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; Pubmed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
 RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Meche R., Mueller M.,
 RA Kretsch M., Delany M., Pulidomenich P., Watson M., Schmidheini T.,
 RA Reichert B., Portelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,

RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weljens I., Voet M., Bastiaens I., Aert R., Deltor E.,
 RA Weltzienegger T., Bothe G., Ransperger U., Hillbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirse W.,
 RA Moollman P., Klein Lankhorst R., Rose M., Hauf J., Koellner P.,
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quill M., Bray Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIlroy R.,
 RA Petlett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
 RA Borkova D., Bioecker H., Scharfe M., Grimm M., Loelner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Farmann B., Grandtath K., Dauner D., Herzl A.,
 RA Neuman S., Argitay A., Vitale D., Liguori R., Piravandi E.,
 RA Massenet O., Quigley F., Clabaud G., Mendelsohn A., Faltner R.,
 RA Schnabel S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheffor F., Cooke R., Berger C., Montfort A., Casachetla E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Reyes A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heljnen L., Schwarz S., Scholler P., Heber S., Francis P., Bleke C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parrell L., Dedha N., Gnoj L., Schultz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Shee P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kallacki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Klerman J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Speth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Sheher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
 RA Chen E., Maria M., Martensen R., McComble W.R.;
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana";
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: MIGHT BE INVOLVED IN THE BIOSYNTHESIS OF STEROIDS
 CC IMPORTANT FOR THE POLAR ELONGATION OF CELLS DURING DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
 CC (Potential).
 CC -1- TISSUE SPECIFICITY: UNOBTAINED.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB008097; BAA37167.1; -
 DR EMBL: AL161589; CAB80304.1; ALT_INIT.
 DR EMBL: Z99708; CAB16850.1; ALT_INIT.
 DR EMBL: AL022141; CAAL8139.1; ALT_SEQ.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; p450.1.
 DR PRINTS: PR00385; p450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme;
 KW Endoplasmic reticulum; Multigene family;
 KW TRANSMEM 4 24
 FT BINDING 463 463 HEME (BY SIMILARITY).
 FT CONFLICT 45 45 F -> L (IN REF. 2).
 SQ SEQUENCE 524 AA: 59389 MW: 550578908BDDF272 CXC64;
 Query Match 32.2%; Score 864; DB 1; Length 524;
 Best Local Similarity 35.4%; Pred. No. 1.3e-44;
 Matches 186; Conservative 96; Mismatches 165; Indels 78; Gaps 10;
 QY 24 LLLLSLSLSLLFLILKRRNKRTRNPPGSGWFLGCTTGYLKPATYATLGDGMOOH 71
 7 LLLLSLSLSLSLLFLILKRRNKRTRNPPGSGWFLGCTTGYLKPATYATLGDGMOOH 71
 Db 7 LLLLSLSLSLSLLFLILKRRNKRTRNPPGSGWFLGCTTGYLKPATYATLGDGMOOH 71

```

DB 30 FLVLTAGILLRPMFLRLRSLKTRKDDDEEDNEKKGMIPNLSGLPVGITLNFAC 89
OY 57 ---KPYATATLGDPMOAHVSKYKIRSNLGEPTIVSADAGLNRETLQNEGRLEFCS 113
DB 90 YSRPVT-----FMDKRSLGKVFERTNIGPPIISTDAEVNKKVYLQHNHGFVAPR 143
OY 114 KSLGGLGKSMVLVGDHMRMSISLNTSHARLTILKDVERTLFLVDSWOONST 173
DB 144 KSTTELGENSIISINGPHOKRLHTLIGAFRLSPHLKDRITRDIEASVLTSLASMAOLPL 203
OY 174 FSNODEAKKFTFMKMHISMNDGPEETEQLKREYTEMKGVASAPLNPGRAYHIALQ 233
DB 204 VHVODELKKATFEILKVLNLSISPG-EDMNLKLEFEETKGLCLIPKPGFRLYKSLK 262
OY 234 SRATILKFERKMERKLDIKEEDDEEEVKT-----EDEAKMSKSDHVRKQRTDDLL 287
DB 263 AKERLIKMKVKKVVEERQVANTTSPANDVYDVLROGDSEKOSQSPDVSQK----- 315
OY 288 GWLKHNSLSTEQILDLILSLFAGHETSSVAIALAIFLQACPKRAVEELREHLEIARA 347
DB 316 -----IYEMMIGETEMPTAMTLAVKFLSDNPVALAKLVEEMEKRR 358
OY 348 KRELSELMWDDYKMDFTQCYINETRLGNVYRFLHRAKLDVRYKGDIPSGKVLVP 407
DB 359 KLELGE-EYKWTYMSLSTFQVNIETLRMANINGWRKALDVEIKGLIPGMCVLA 417
OY 408 VISAVHLNDRSYDQPNLFNPMRMOQONNGASSGGSFSTWGNVYMPFGGPRCAGSEL 467
DB 418 SFSVHMDIEDIYNPQFDPWRMDRINGSANSSIC-----FTFPGGQRLCPLEL 468
OY 468 AKLEMAVFIHHLVKFENWELAEDDQPAFPFVDPNGLPPIRVSTI 512
DB 469 SKLEISIFLHLVTRYSW-FAEDEIVSFPTVKMKRRLPIRVATV 512

RESULT 3
CP85_LYCES STANDARD: PRT: 464 AA.
ID CP85_LYCES 043147;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
CYTOCHROME P450 85 (EC 1.14.-.-) (Dwarf protein).
GN CYP85 OR D.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. GCR758;
RX MEDLINE=6626705; PubMed=8672892;
RA Bishop G.J., Harrison K., Jones J.G.D.;
  "The tomato Dwarf gene isolated by heterologous transposon tagging
  encodes the first member of a new cytochrome P450 family.";
  Plant Cell 8:959-969(1996).
RT -1. SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
RL
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U54770; AAB17070.1;
DR Interp: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.

```

```

KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BINDING 414 414 HEME (BY SIMILARITY).
SQ SEQUENCE 464 AA; 53706 MW; D2B31AABVBI4E94 CRC64;

Query Match 28.4%; Score 761.5; DB 1; Length 464;
Best Local Similarity 33.4%; Pred. No. 1,4e-38;
Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

OY 12 LLLPRLSLLEFLILKRRN--RKTFRNLPFGSGMFLGEGTGYLKPATATLG-DFM 68
DB 5 LIFLSFEGICIFRFTALLRNQVYKNLPFGMGMPRLGGETTERFLK-----LGRSPM 58
OY 69 QOHVSKYKIRSNLGEPTIVSADAGLNRETLQNEGRLEFCSYPSRISGIGLGRMSDLV 128
DB 59 KQORARYGEFSHILCCPIVSDSELNRYLIIVNEAKGLVPGQGMIDILGRCNIAAV 118
OY 129 VGDMHMRMSISLNTSHARLTILKDVERTLFLVDSWOONSTIFSAODEAKFTTNLM 188
DB 119 NSAHKYMKGALLSLISPTMIROQLPRIDEFMRSHLTNW-DNKVIDIOERTNMAFLSS 177
OY 189 AKHISMNDGPEET---EOLKREYTEMKGVASAPLNPGRAYHAKALOSRATILKFERK 245
DB 178 LKQI-----AGLESTSLAQEFMSFEFNVLGTLSPILNPTNHYRGQAKIYVNLRTL 233
OY 246 MEERKLDIKEEDDEEEVKTDEEAEMSKSDHVRKQRTDDLLGWLKHS---NLSTEQI 301
DB 234 IEERR-----ASKEIQHMDLVLMNEEATRFKLTODEM 266
OY 302 LDILSLFAGHETSSVAIALAIFLQACPKRAVEELREHLEIARAKELGESLMDDY 361
DB 267 IDLITLYSGYETVSTSMNAVKYLLDHPVLEELKEHMAIIEKKRP-EDIDIDNY 324
OY 362 KMDFTQCYINETRLGNVYRFLHRAKLDVRYKGDIPSGKVLPIYSAVHLNDRSYDQ 421
DB 325 RSMRFTRAVLETRLATIYNGVLRKTTQDMKINGIILPKGRKRYVYTRELANDPRXPD 384
OY 422 PNLFNPMRMOQONNGASSGGSFSTWGNVYMPFGGPRCAGSELAKLEMAVFIHHLVL 481
DB 385 PYSEFPMRMDKS-----LEHQNSFLVGGGTRQCPGKEKLVAAISTFLHVFVT 433
OY 482 KFNWELAEDDQPAFPFVDPNGLPPIRVSTI 510
DB 434 KYRWEELGDKLMPFEVAPNGLPPIRVSTI 462

RESULT 4
C883_ARATH STANDARD: PRT: 490 AA.
ID C883_ARATH 023051;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
CYTOCHROME P450 88A3 (EC 1.14.-.-).
GN CYP88A3 OR AT1G05160 OR YUP8H12.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae;
  eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Columbia; PubMed=11130712;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
  White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
  Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
  Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
  Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
  RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
  Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
  RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kvan A., Lam B.,
  Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
  Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,

```

RA Miltcher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shin P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utecht T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.", 816-820(2000).
RL Nature 408:816-820(2000).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AC000098; AAB71462.1; -
DR HSRP: P14779; IUP2.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT TRANSMEM
FT BINDING 439 439 HEME (BY SIMILARITY).
SQ SEQUENCE 490 AA; 56409 MW; 7FD8CD7A8864D298 CRC64;
Query Match 21.6%; Score 579.5; DB 1; Length 490;
Best Local Similarity 28.6%; Pred. No. 1.1e-27;
Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;
OY 34 KTRFNLPGKSGMPFLGEGTIGYKLPYATTLGDFMOQHSKYGK--IYRSNLFGEPTIVS 91
DB 40 EHRHLPDGLDGLMPLGNLSPFLRAFKTSDDPSFTRLIKRGPGIYVAHNFQNPSTIV 99
OY 92 ADAGLNRFLQNEGRLEFECSTYRSIGGILGKMSLVLDGMDRMSISLNFSLHARLT 151
DB 100 TTSIDCRRLTDD--AFKRGMTSMELGKRSFVGISFEHKKRLRLTAALVNGHEALS 158
OY 152 ILKQVRRHTLVLDSDMOONSIFSAODEAKKFTFNLMKHMISMDPGEEETQLKKEYT 211
DB 159 TITPIEEVITVLDKTKMGFEFLTHIKRLTFRIT--MTIFLSESEVMDALEEETA 217
OY 212 FMKGVVAPLNPCTAYHKAQSRATILKFERKMEERKLDIKEEDOEVEEYKTEDEAM 271
DB 218 LNYGVRAVAVNPFGAYHRAALKAKTLVAAPQSYTER--NQRQNILSKKMDL 271
OY 272 SKSDHVRKQRTDDLLGWVLKHSNSTEODILILSLFAGHETSSVAIALAIFLQACP 331
DB 272 DMLNWK-----DEK-----GKTLDEDEIIDLMLYLNAGHESGHTIMATVLEQEP 320
OY 332 KAVEELREHLEIARAKKELGESLMDYKMDFTQCVINTRLGNVRLHRAKLD 391
DB 321 EYLQAKADEQEMILKSRP--GQKGLSLKETRMELFSGVDETLLVITFSLTAPEAKTD 379
OY 392 VNYKGYDIPSGKVLPIVISAHLDSRYDQPLFNPMWQOONNAGSSGSGSFTWGN 451
DB 380 VEMNGLYLPKGMKVLWVRDVAIDPEVFPDPKFPAPR--DNG-----FVPKACA 428
OY 452 YMPFGGPRLCAGSELAKEMAVTTHLVLFKFMELADDDQFAP 496
DB 429 FLPEFAGSHLCPGNDLAKLEISIFLHFLFKYQVRSNDECPVM 473
RESULT 5
CP26_BRAKE STANDARD: PRF; 492 AA.
AC P79739;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytochrome P450 26 (EC 1.14.11.-) (Retinoic acid-metabolizing
DE cytochrome) (P450RA1) (Retinoic acid 4-hydroxylase).
GN CYP26.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Euryarchaea; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094702; PubMed=8939936;
RA White J.A., Guo Y.-D., Baetz K., Beckett-Jones B., Bonasoro J.,
RA Hsu K.B., Dilworth F.J., Jones G., Petkovich M.;
RT "Identification of the retinoic acid-inducible all-trans-retinoic
RT acid 4-hydroxylase".
RL J. Biol. Chem. 271:29922-29927(1996).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOLIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF 4-HYDROXYLATION, HYDROXYLATION,
CC RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED FORMS OF RA,
CC INCLUDING 4-OH-RA AND 4-OXO-RA.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- INDUCTION: BY RETINOIC ACID.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U68234; AAC60045.1; -
DR ZFIN: ZDB-GENE-990415-44; CYP26.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum
FT BINDING 438 438 HEME (BY SIMILARITY)
SQ SEQUENCE 492 AA; 56281 MW; F0471435B2F30509 CRC64;
Query Match 18.8%; Score 503; DB 1; Length 492;
Best Local Similarity 27.6%; Pred. No. 3.8e-23;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;
OY 7 HTLLPLLLPSLLSLFL-----ILKRRNRKTRNLPKSGMPFLGEGTIGYKLP 58
DB 4 YLMTVFTCTIYLPVLLFLAAVKLWMLIRRYDNCSPPLPGTMGLPFTIGTQLI-- 61
OY 59 YVATLDPFMOOHVSKYGYKTVRSNLFGEPTIVSADAGLNRLFLQNEGRLEFECSTYRSIG 118
DB 62 ---LDRKFLRKRRKROKIGCTITFLFGNFTVAGADNVRLQLGHEHLVSYQWAPASVPT 118
OY 119 ILGKMSLVLDGMDRD-----MRSISLNFSLHARLTILKQVRRHTLVLDSDV--QONS 172
DB 119 ILGSDTLNVHVGQHKNNKKAIRAFSDALEH-----YIPVIOQEVKSAQEMLOKDS 172
OY 173 IFSADDEAKKFTFNLMKHMISMDPGEEET--EQLKKEYVTFMKGVVAPLNPCTAYH 230
DB 173 CVLVPEPKKILMFRAMILLGFEEQKLTQDEQLEVAFEEMIKMLFSLPIDVPSGLYR 232
OY 231 ALQSRATILKFERKMEERKLDIKEEDOEVEEYKTEDEAKMSKSHVAKQRTDDLLGW 290
DB 233 GLRAR-----NPLHSKIEENIRKKIQQDDNENQKYLQDL--INSRSDE----- 279
OY 291 LKHSNLSTEODILILSLFAGHETSSVAIALAIFLQACPAVELEHREHLEIARAKE 350
DB 280 ---PFSIQAKKEATELLFGHETTAATSTATSLVNLGLNTEVQAKVE---VQEKVE 331

QY 351 LG-----ESELNMDYKKMDFOCVINEELRLGNVRFELHKKALKDVRVRYGYDIPSGWKVL 406
 Db 332 MGRYTGKGLSMELLDQLKTYGCIKETTILINPVPYGGFRVALKTYPELKGIOYIPKGMNYI 391
 QY 407 PVISAVHLNDSRYDQNLFPNPMWMOOONNGASSSGSFSTGMNNMPPRGSGPRLCAGSE 466
 Db 392 YSICDHFVADVFNKNEEOPREFM-----SKGLEDSRF--NIIPEGGSRCMKVGE 442
 QY 467 LAKLENAVITLHLVLFKNELAEDDO-----PPAPFVDFP 502
 Db 443 FAKVLKIFLVELTCHCNWILSNGPPTMKGTPTIYFVDNL 483

RESULT 6

C881_MAIZE STANDARD: PRT; 519 AA.
 AC 043246;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome P450 88A1 (EC 1.14.-.-) (DMAKF3 protein).
 GN CY88A1 OR D3
 OS Zea mays (Maize)
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. B73;
 RX MEDLINE=96004534; PubMed=7549486;
 RA Winkler R.G., Helentjaris T.;
 RT "The maize Dwarf3 gene encodes a cytochrome P450-mediated early step
 in gibberellin biosynthesis.";
 RL Plant Cell 7:1307-1317(1995).
 CC -1- PARTWAY: Gibberellin biosynthesis; early step.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, DEVELOPING LEAVES, THE
 CC -1- VEGETATIVE MERISTEM, AND SUSPENSION CULTURE CELLS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32579; AAC49067.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
 FT TRANSMEM 1 21 PORENITAL.
 FT BINDING 466 466 HEME (BY SIMILARITY).
 SO SEQUENCE 519 AA; 57906 MW; 0F8977A024316D95 CRC64;

Query Match 18.3%; Score 491.5; DB 1; Length 519;
 Best Local Similarity 28.2%; Pred. No. 2e-22;
 Matches 129; Conservative 91; Mismatches 189; Indels 49; Gaps 11;

QY 36 RFNLPKRSQMPFLGETITGLKPTATTLADPEMOQVHXYGK--IYRSNLFGPTI-VSA 92
 Db 69 RARLPPEKSMPLVGGMMALRAFKSPDAFASFRGRGVRSGVRSFMSPTVLVYT 128
 QY 93 DAGINRFLILOGRFLFCSPYRISGILGKSMVLVVGDMHRRMSISLNFSLHARLRT 152
 Db 129 AEGQKVLMDDA--FYTGPKATVALVGRSVVAMPYDHRIRIKILTAAPINGPDLATG 186
 QY 153 LLDQVHRTFLVDSW-QQNSIFSADKAKFTFNLMAKHIMSDPGEEETDLKKEYT 211
 Db 187 YLPFDRTVTSIRAWADHGSGVEFLTELARMPFKIIIV-QIFGAGDAQATTRALERSTYE 245

QY 212 FMKGVASAPLNPSTAYHAKLOSRAITIKFTIERKMEERKLIDK-----EDQEEVEYKT 265
 Db 246 LNYGRKAAALNIPGRATYALARRRVAIVLOGVLDERRARRAKVSGGGVMDRLIEA 305
 QY 266 EDEAEKMSKSDHYRKORTDDLLGLWLNKSNLSTEOILILISLIFAGHETSVAIALAIF 325
 Db 306 QDE-----RGRHLLDD-----ELIVLVWYINAGHESGHTTMATV 342
 QY 326 FLQACPRAVEELREHLEIARAKKELGESELNMDYKKMDFOCVINEELRLGNVRFELH 385
 Db 343 FLQENPDMFARAKAEQEAIMRSIPS-SQRLGLLDFRMEYLSQVIDETLAVNISFVSF 401
 QY 386 RKALKDVRKGYDIPSGWKVLPVISAVHLNDSRYDQNLFPNPMWMOOONNGASSSGSF 445
 Db 402 RQATRDVFNKTYLIPKKGKVLWTRVHMDPOVYDPDKFPSPHWE-----GHSPRAGTF 456
 QY 446 STWGNMYPFGSGPRLCAGSELAKLENAVITLHLVLF 483
 Db 457 -----LAFGLGARLCPCGNDLAKLEISVLFHFLDGY 487

RESULT 7

C120_SYNY3 STANDARD: PRT; 444 AA.
 AC Q35990;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative cytochrome P450 120 (EC 1.14.-.-).
 GN CYP120 OR CYP OR SUR0574.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 RA Sugita M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 RL DNA Res. 2:153-166(1995).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=97061201; PubMed=8905231;
 CC Kaneo T., Sato S., Kotani H., Tanaka A., Asanizu E., Nakamura Y.,
 CC Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
 CC Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,
 CC Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 CC Tabata S.;
 CC "Sequence analysis of the genome of the unicellular cyanobacterium
 CC Synechocystis sp. strain PCC6803. II. Sequence determination of the
 CC entire genome and assignment of potential protein-coding regions.";
 CC DNA Res. 3:109-136(1996).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 DR EMBL: D64003; BAA10496.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
 KW Complete proteome.
 FT BINDING 391 391 HEME (BY SIMILARITY).
 SO SEQUENCE 444 AA; 50578 MW; 8F62A9EBD3B54BDC CRC64;

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

Query Match 17.9%; Score 479.5; DB 1; Length 444;
 Best Local Similarity 29.7%; Fred. No. 8.5e-22;
 Matches 137; Conservative 88; Mismatches 167; Indels 69; Gaps 13;

DB 13 IPPGDFGLPMLGETINFLND-----GDPGRKROQFGIFTRLFGRKAVITISGALANR 66
 39 IPPGSGMFPLEETIGYLPYATATGLDPMOQHSKYGKIRYSNLEGPETYSADAGLNR 98
 OY 13 IPPGSGMFPLEETIGYLPYATATGLDPMOQHSKYGKIRYSNLEGPETYSADAGLNR 98
 OY 99 FTLQNEGRLEFCSYPSRISIGILGKMSMLVVGDMHRMDSISLNFSLHART-RTT-LLK 155
 DB 67 FLFTKEQEFQATWPLSTFILLGPNALATQMGCEIHRSRKI---LYQAFLPRTIDSYLP 122
 OY 156 DYERHTLPVLDWQONSIFSADEAKFTFNMAKHISMNDGEE--ETEQKKEYVTFM 213
 DB 123 KMDGIYGVGLQWGRKANEYIWPOLRMFTEDVAATLFM---GEVYSQNPOLFPEFTYI 178
 OY 214 KGVASAPLPLPCTAYHAKLQSRATILKFTERMEERKLDIKREDDQEEBEVKTEDAEKSK 273
 DB 179 QGLFSLPRLPRTLFGKSGSRARALL-----AELEK 209
 OY 274 SDHVRKQR--TDDLLGWVL-----KHSNLSTEQILDLILSLFAGHETSSVALALATFP 326
 DB 210 IIRAKQOQPPSEEDALGILLARDNNOPSLPELKDQILLILFAGHETLFSALSPCLL 269
 OY 327 LQACRAVELBEHLERARAKKEGSESLNMDYKKKDFTCVINEIRLGNVYRFLHR 386
 DB 270 LGQHSDIRRVRQOE-----QNLQLQSQ--ELTAEKTKMPEYLDQVQLEVRLRLPGVGGR 323
 OY 387 KALKDVRKGYDIPSGMKVLPVTSVHLDNSRYDQNPENFPMRMOQNNAGSSGSGSPS 446
 DB 324 ELIQDCQOGFFPKGWLVSYSIOTHADPDLPPPEKDFPRFTPD---GSATINPPFA 380
 OY 447 TWGNMYPGGGPRLCAGSELAKLEMAYFIHHLVLEKFMNEL 487
 DB 381 -----HVEFGGGLRCLGKEFARLEKLFATRLIQOFPDWTL 416

RESULT 8
 CP26_MOUSE STANDARD; PRT; 497 AA.
 ID CP26_MOUSE
 AC 055127; O9RI1F4;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoic acid-metabolizing
 cytochrome) (P450RA1) (Retinoic acid 4-hydroxylase).
 GN CYP26A1 OR CYP26 OR P450RA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H;
 RX MEDLINE=97392446; Pubmed=9250660;
 RA Fujii H., Sato T., Kaneko S., Gotoh O., Fujii-Kuriyama Y., Osawa K.,
 Kato S., Hamada H.;
 RT "Metabolic inactivation of retinoic acid by a novel P450
 RT differentially expressed in developing mouse embryos.";
 RL EMO J. 16:4163-4173(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9811312; Pubmed=9442090;
 RA Abu-Abed S.S., Beckett B.R., Chiba H., Chithalen J.V., Jones G.,
 Metzger D., Chambon P., Petkovich M.;
 RT "Mouse P450RA1 (CYP26) expression and retinoic acid-inducible retinoic
 RT acid metabolism in F9 cells are regulated by retinoic acid receptor
 RT gamma and retinoid X receptor alpha.";
 RL J. Biol. Chem. 273:2409-2415(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Tooth;
 RX MEDLINE=20515038; Pubmed=11063033;

RA Paule C.T., Paule M.L., Snead M.L.;
 RT "Identification of tufellin- and amelogenin-interacting proteins using
 RT the yeast two-hybrid system.";
 RL Connect. Tissue Res. 38:257-267(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Plays a key role in retinoic acid metabolism. Acts on
 CC retinoids, including all-trans-retinoic acid (RA) and its
 CC stereoisomer 9-cis-RA. Capable of both 4-hydroxylation and 18-
 CC hydroxylation. Responsible for generation of several hydroxylated
 CC forms of RA, including 4-OH-RA, 4-oxo-RA, and 18-OH-RA.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; Y12657; CAA73206.1; -;
 DR EMBL; AF115769; AAD17217.1; -;
 DR EMBL; BC012673; AAH12673.1; -;
 DR HSSP; P14779; 1JPZ.
 DR MGI; MGI:1096359; Cyp26a1.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
 KM Endoplasmic reticulum.
 FT BINDING 442 442 HEME (POTENTIAL).
 FT CONFLICT 9 9 S -> T (IN REF. 3).
 FT CONFLICT 154 154 L -> P (IN REF. 4).
 FT CONFLICT 356 356 I -> T (IN REF. 4).
 FT CONFLICT 492 492 Y -> H (IN REF. 4).
 SQ SEQUENCE 497 AA: 56177 MW: 33807D7C29134471 CAC64;

Query Match 17.2%; Score 461; DB 1; Length 497;
 Best Local Similarity 28.4%; Fred. No. 1.2e-20;
 Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

OY 10 LPALLPLSP-----LSLLFLILLK-----RRRKTRFNLPKSGMFPLEETIGYLIK 57
 DB 3 LPALLASALCTFVLPPLFLFLAALKMDLYCVSSRDSICALPLPPTMGFPFFGETL--- 58
 OY 58 PYTATTLDFMOQHSKYGKIRYSNLEGPETYSADAGLNFRTLQNEGRLEFCSYPSRISG 117
 DB 59 -QMVQRRKFTLQMKRRKKGFTIKHLFGFRYRVAGADNVRKILLGHRILVSVMHPASVR 117
 OY 118 GILGWSMLVVGDMHRMDSISLNFSLHARTLTLQDVERHTLFLVDSWQO--NSIFS 175
 DB 118 TLTGAGCLSNLHDSHKORKKVIQAFSREALQCVLV-IAEVSCLQEWMLSCGERGLL 176
 OY 176 AODEAKKFTFNMAKHISMNDGEE--ETEQKKEYVTFPKGVASAPLPLPGATYHKA 231
 DB 177 YYPEVKRLMFRIAMRILLCGEPGAGGEDQQLVAEPEEMTRNLFSPIDVPFGLYRG 236
 OY 232 LOSRVATILKFIERKME--RKLDIKEED--QEEBEVTEDEAKSKSDHVAKORTDDLL 286
 DB 237 VKARULIARIEFNIRAIRRLQATERPBGCCDADQLILHSWE-----RGEYLDMO- 288
 OY 287 LGWVLRKSNLSTEQILDLILSLFAGHETSSVALALATFLOACRAVELBEHLERAR 346
 DB 289 ---ALKQS---STE-----LLFGHETTSAAASLITVGLVPHVQLKAVEELKSKGL 335
 OY 347 AKKELGSELNMDYKKKDFTCVINEIRLGNVYRFLHRKALKDVRKGYDIPSGMKVYL 406

```

Db 336 LCSNMONKIDMETEOLKICVYKELRLNPPVGGFRALTKTFELNGIOTKGMVYI 395
Qy 407 PVISAVHLDSRYDQPLFNPMWMOQONNGASSSGSFTWGNMYPFGGPRICAGSE 466
Db 396 YSICDTHDVAIDFTNKEEFNDRIYVHPEDASRFS-----FIPFGGLASCYKE 446
Qy 467 LAKIEMAVFTLHHLVTKFNWELAE-----DDQFAFPPVDFP 502
Db 447 FAKILKIFVVELARHCDQMLNGPPTMKTSPTVYVNDLP 487

RESULT 9
C26A_HUMAN STANDARD; PRT; 512 AA.
ID C26A_HUMAN
AC Q9NR63; Q9NP41;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Cytochrome P450 26A2 (EC 1.14.-.-) (P450RA1-2) (Retinoic-acid
  metabolizing cytochrome).
GN CYP26A2 OR P450RA12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=20300913; PubMed=10823918;
RA White J.A., Ramshaw H., Taimi M., Stangle W., Zhang A., Everingham S.,
RA Creighton S., Tam S.-P., Jones G., Petkovich M.;
RT "Identification of the human cytochrome P450, P450RA1-2, which is
RT predominantly expressed in the adult cerebellum and is responsible for
RT all-trans-retinoic acid metabolism."
RT Proc. Natl. Acad. Sci. U.S.A. 97:6403-6408(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Bourne S., Bauer C., Pape K., Jones T.;
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays a key role in retinoic acid metabolism. Involved
CC in the specific inactivation of all-trans-RA. Responsible for
CC generation of several hydroxylated forms of RA, including 4-OH-RA,
CC 4-oxo-RA, and 18-OH-RA.
CC -1- ENZYME REGULATION: Has a preferred activity towards the following
CC substrates: all-trans-RA > 9-cis-RA > 13-cis-RA.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC similarity).
CC -1- TISSUE SPECIFICITY: Highly expressed in brain, particularly in the
CC cerebellum and pons.
CC -1- INDUCTION: By retinoic acids (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF252297; AAF76003.1; -
CC EMBL: AC007002; AAF65576.1; -
CC MIM: 605207; -
CC InterPro: IPR001128; Cytochrome_P450.
CC Pfam: PF00067; P450_2.
CC PRINTS: PR00385; P450.
CC PROSITE: PS00086; CYTOCHROME_P450; 1.
CC Oxidoreductase; Monooxygenase; Membrane; Heme; Mitosome;
CC Endoplasmic reticulum.
CC BINDING 441 441 HEME (POTENTIAL).
CC CONFLICT 68 68 O -> OKCTAREPVMPO (IN REF. 2).
CC SEQUENCE 512 AA; 57512 MW; A06D1D9944E6726F CRC64;

```

```

Query Match 16.6%; Score 446; DB 1; Length 512;
Best Local Similarity 26.9%; Pred. No. 9.9e-20;
Matches 132; Conservative 86; Mismatches 219; Indels 54; Gaps 10;

Qy 14 LPLSLSLFLFLILKLR-----RRKTRFNLPEKSGMFPPLGELTGLKPTATTL 64
Db 15 LAACLVSVTLTLLAVSQQLRMATRTDRKSKLPRKSGMFPPLGELTGLKPTATTL 71
Qy 65 GDFMOQHRSKYKGIYRSLNLFCEPTIVSADAGINRPILONEGRLEFCSYPRISGILGKWS 124
Db 72 --FOSSRREKGNVFKTLLGRPLIRYTGAEVNRKILNGEHLIVSTPMRSTRMLIGNT 129
Qy 125 MLVLVGDNRHMRISLNLFLSHARLTLLDVERHTLFLVDSWQON-SIFSADDEARKE 183
Db 130 VSNISGDIHRKRKRKVFYSFHEALESYLPK-IGLVIDTLRAMSGHDEAINVQEAQKL 188
Qy 184 TPNLMARKIIMSMDREETEDLKKRYVFMKGVSAPRLNLEGTAHKLQSRATILKTE 243
Db 189 TFRMAIRVLLFSPISPEEDLGHLEFYVQFVDNVFSLPVDLPSPGRRIGIARQILQKLE 248
Qy 244 RKMEERKIDKEEDQEEVEKTEDEAEKSDHVRKQRTDDLLGWLKHSNLSTEOIID 303
Db 249 KAIREK-----LQYQGRKYDALDLILLESSKEHKE-----MTWQELKD 288
Qy 304 LILSLFAGHETSSVAIALAIFLQACPRAYEELREHLEZARAKKEG-----ESELN 357
Db 289 GTLELIFAAVYVTTASASTSLIMQLKHPVLEKLNDE---LRANGILHSGGCEPGLTR 344
Qy 358 WDDKKMDFTQCVINETRLGNVAVFELIRKALKDVRKYGYDIPSGKVLPAVAVHLDS 417
Db 345 LDTLSGLLYLCCILKEVRLFTPISGYRTRVLOFELDFQIPKWSYSTRDHDPAP 404
Qy 418 RYDQNLNFMWMOQONNGASSSGSFTWGNMYPFGGPRICAGSELAKLEMAVPIH 477
Db 405 VFQDVNVDFDPDFSG---ARSEDKDGR-----HILPFGGVRTGLGKHLAKFLKVLAV 456
Qy 478 HLVLKFNWELA 488
Db 457 ELASTSRELA 467

RESULT 10
CP26_XENLA STANDARD; PRT; 492 AA.
ID CP26_XENLA
AC O93323;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoic acid degrading enzyme
  CYP26) (XCP26) (Retinoic acid converting enzyme) (RACE).
GN CYP26A1 OR CYP26.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99077803; PubMed=9857192;
RA Hollmann T., Chen Y., Grunz H., Pleier T.;
RT "Regionalized metabolic activity establishes boundaries of retinoic
RT acid signaling."
RL EMOB J. 17:7361-7372(1998).
CC -1- FUNCTION: Plays a key role in retinoic acid metabolism (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC similarity).
CC -1- INDUCTION: By retinoic acid.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL: AF057566; AAC25158.1; -
InterPro: IPR001128; Cytochrome_P450.
Pfam: PF00067; P450; 1.
PRINTS: PR00385; P450.
PROSITE: PS00086; CYTOCHROME_P450; 1.
Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
Endoplasmic reticulum.
KW BINDING 437 HEME (POTENTIAL).
FT SEQUENCE 492 AA; 55459 MW; D1DAB7651BF2D3E CRC64;
SQ

Query Match 16.6%; Score 444.5; DB 1; Length 492;
Best Local Similarity 28.7%; Pred. No. 1.2e-19;
Matches 149; Conservative 85; Mismatches 222; Indels 63; Gaps 17;

QY 7 HTLLPLLLPSLLSLFLT-----ILKRRNRKTRNLPKSGMPFLGTTGLKLP 58
DB 4 YTLTSLALCTLALPRLTLTAALMEVYCLRRKADACANPLPGTGLPFGFTGLMVLQ 63
QY 59 YVATITGDMQOQVSKYKTIYSNLFGEPTIVSADAGLAFIQLNGRLPEGSPISIG 118
DB 64 RRR-----FLQYKRSQYGRITKTHLGSPTRVATGAEVNRQIIMGHRKIVSVMPASVET 118
QY 119 ILGKMSLVLVGDHMRDMSISLNFSLHARLTILKDVERTILFLVDSMOONS---IF 174
DB 119 ILGAGLSLMDHNEHKYTKVIAQAFSRALAN-YYPQHEEYRCVNLMLDSGCVLVY 177
QY 175 SAODEKATFTFLMAKHIMSDP---GEEETDLKREYVTFMGVSAPLNPGTAYHKA 231
DB 178 PA---IKRMKFRIMRLILGCDPQRMDRQETLLLEAFEMSRLFLSDIVPFGSLYRG 234
QY 232 LOSRATILFIERKMERLIDKEDEEYKTEDEAMSKSDHAKRTDDLLGWL 291
DB 235 LRRR---NLHQAQIE---NTEKLRQRPDEHKADQL-LIDYSR-RNGEPIVQLAL 284
QY 292 KHSNLTEDILDLISLIFAGHETSSVAIALAIFFLQACPKAVEELREHLE---TARAK 348
DB 285 KSS--ATE-----LFGGHGTTASATSLTSLFLAKDLKDLKRAKE-LLEQGLSLTK 333
QY 349 KELGESELNWDYKKMDFQCYINETLRIGNVYRFLHRAKLDVRYKYDIDSGKVLVY 408
DB 334 PE-EKKELEISIEVLQOLKTSYKIVKTELRLSPVAGGFVALKTFVNGYQIRKGMVYIS 392
QY 409 ISAVHLDNSRYQPNLFNPMWQOQNGNASSSGSSFSFSGWNNYMPFGGPRICASELA 468
DB 393 IADTGHGADLFPTDKFNDFRFLTLPRDSSR-----RGIIPFGGVCRCIGKEPA 443
QY 469 KLEMAVFIHHLVLFKFMEL-----AEDDPFAFPFVDFP 502
DB 444 KILKVFVVELCRNCMDWELNLSPAWTSPITICPVNDLP 482

RESULT 11

CP26_CHICK STANDARD; PRT; 492 AA.
AC 09PUB4: 09PUG2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoid acid degrading enzyme CYP26).
GN CYP26A1 OR CYP26.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic spinal cord;
RX MEDLINE=20057772; PubMed=10588879;
RA Swidell E.C., Thaller C., Sockanathan S., Petkovich M., Jessell T.M., Etchells G.
RT "Complementary domains of retinoid acid production and degradation in the early chick embryo."
RT Dev. Biol. 216:282-296(1999).
RN [2]
RP SEQUENCE OF 368-474 FROM N.A.
RA Martinez-Ceballos E., Burdall C.A.;
RL Submitted (SEP1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays a key role in retinoid acid metabolism. Appears to be involved in the specific inactivation of all-trans-RA.
CC Responsible for generation of several hydroxylated forms of RA, including 4-OH-RA, 4-oxo-RA, and 18-OH-RA.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- DEVELOPMENTAL STAGE: Expressed at stage 4 in the ectoderm, stage 5-7 in the nascent notochord and at stage 7 its expression decreases in the anterior part of the embryo. From stage 7-10 its expression is restricted to the dorsal folds of the neural tube and to rhombomere 2. At stage 10, it is expressed in the lateral plate endoderm and in the tail bud and by stage 11/12 it disappears in the neural tube, followed by a confined expression at stage 12 to dorsal neural tube and at stage 15 an increasing expression in the ectoderm.
CC -1- INDUCTION: By retinoid acid.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL: AF199462; AAF09250.1; -
EMBL: AF185266; AAD56546.1; -
InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 2.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 437 437 HEME (POTENTIAL).
FT CONFLICT 371 371 I -> V (IN REF. 2).
FT CONFLICT 472 474 GPI -> SPT (IN REF. 2).
SQ SEQUENCE 492 AA; 55264 MW; 7F28B72E75C2322FB CRC64;

Query Match 16.4%; Score 439.5; DB 1; Length 492;
Best Local Similarity 26.3%; Pred. No. 2.3e-19;
Matches 141; Conservative 84; Mismatches 210; Indels 101; Gaps 15;

QY 9 LRLPLLLPSLLSLFLILK-----RRNKRTRFNLPKSGMPFLGTTGLKPYT 60
DB 6 LVASALCTFLPLLLFLAVALMDLYCASGRDPCSLPLPGTGLPFGFTGLT-----QM 60
QY 61 ATTIGDMQOQVSKYKTIYSNLFGEPTIVSADAGLAFIQLNGRLPEGSPISIGTIL 120
DB 61 VLQRRKFLDMKRRKRTGYITKTHLFGRPYRWAGAEVNRHILGHNLYSVQMPASVRLTL 120
QY 121 GRKMSLVLVGDHMRDMSISLNFSLHARLTILKDVERTILFLVDSMOONSIFSAQ--- 177
DB 121 GSGCLSNLHNGCHKHKKVIMQAFS-----RDALGHVYVLD--EVSACLAQWIG 169
QY 178 -----DEAKKTFPFLMAKHIMSDPGE---EETDLKREYVTFMGVSAPLNPGT 226
DB 170 AGPCLLYVEVNRKLMRIMRILGFPQASPDGQQLVEAFEEIRIMLFSLIDVPS 229
QY 227 AVHKLQSRATILKFERKMERKLDKEDEEYKTEDEAMSKSDHAKRTDDLL 286
DB 230 GLYGLRRR---NTHAKIEE-----NIRAKMARKE---PEGYKDA 265

```

QY 287 LGWLKHS-----NLSTEQIIDLILSLFAGHETSVAIALAIFPLQACPAVEELREEH 341
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 266 LQLMHEHTQNGEQLNMQELKESATELLEFGHETTASATSLIFLGI-----HH 315
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 342 LEIARAKEL-----GESELMNDYKKMDPTOCVINEETRLGNVAFRLHRAKALD 391
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 316 DVLQKVRKRELDQKLGSLSPNOEKQNMFEQLQKTYGCVIETRLSPVPGGFRILAKT 375
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 392 VRYGYDIPSGWKVLPVISAHLNDSRYDQPLNPNWMOQONGASSSGSSTWGN 451
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 376 LELNGYQIPKGMVNYISICDTHDVADELFTDKDEFNPDRFM-----SPSPEDSRRS 426
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 452 YMPGGGRCAGSELKLEMAVFIHHLVLFKNELAEADQ-----PPAPFPVDP 502
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 427 FIPGGGLRSCVGRKFAKVLKIFTVELARSCDQLNGPPTMKTGPIYVDNLP 482
| : : : : | : : : : | : : : : | : : : : | : : : : |

RESULT 12
CP26_HUMAN
ID CP26_HUMAN STANDARD; PRT; 497 AA.
AC 043174;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 26 (EC 1.14.--) (Retinoic acid-metabolizing
DE cytochrome) (P450RA1) (hp450RA1) (Retinoic acid 4-hydroxylase).
GN CYP26A1 OR CYP26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373542; PubMed=9228017;
RA White J.A., Beckett-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,
RA Jones G., Petkovich M.;
RT "DNA cloning of human retinoic acid-metabolizing enzyme (hp450RA1)
RT identifies a novel family of cytochromes P450."
RL J. Biol. Chem. 272:18538-18541(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98380037; PubMed=9716180;
RA Sonnenfeld E., van den Brink C.E., van der Leede B.M., Schulkes R.K.,
RA Petkovich M., van der Burg B., van der Saag P.T.;
RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for
RT all-trans-RA and can be induced through RA receptors in human breast
RT and colon carcinoma cells."
RL Cell Growth Differ. 9:629-637(1998).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=99045433; PubMed=9826557;
RA Trofimova-Griffin M.E., Juchau M.R.;
RT "Expression of cytochrome P450RA1 (CYP26) in human fetal hepatic and
RT cephalic tissues."
RL Biochem. Biophys. Res. Commun. 252:487-491(1998).
CC -!- FUNCTION: Plays a key role in retinoic acid metabolism. Acts on
CC retinoids, including all-trans-retinoic acid (RA) and its
CC stereoisomer 9-cis-RA. Capable of both 4-hydroxylation and 18-
CC hydroxylation. Responsible for generation of several hydroxylated
CC forms of RA, including 4-OH-RA, 4-oxo-RA, AND 18-OH-RA.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -!- TISSUE SPECIFICITY: Highest levels in adult liver, heart,
CC pituitary gland, adrenal gland, placenta and regions of the brain.
CC -!- INDUCTION: By retinoic acid.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF005418; AAB88881.1; -.
DR HSSP: P14779; 1upz.
DR Genew: HGNC:2603; CYP26A1.
DR MIM: 602239; -.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxioreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56162 MW; EAB684B24B2EAB3 CRC64;

Query Match 16.3%; Score 438; DB 1; Length 497;
Best Local Similarity 27.0%; Pred. No. 2.9e-11;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 10 LPLLLPSL-----LSLLFLILK-----RRNRKTRFNLPPGSGMPAGETIGYK 57
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 3 LPALLASALCTFVLPILFLFAIKLMDLYCSGDRSCALPLPGTMGFFFGETL----- 58
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 58 PYTATLGDPEQOQVSKYKTIYRSNLGEPPIVSADAGLNFTIQNEGRLFECSYPSIG 117
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 59 -QWYLQRRKFLQMKRRKGYFTKTHLFGRPYRWGADNVRILLGDRLVSVHPASVR 117
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 118 GILKWSMLVYVGDHMRMSISNLSFLSHARL---TLLKDVREHILFVDSMQ--NS 172
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 118 TILDSGCLSNLHDSHKQKRVYKRAFSREALDECVPIVTEVSS--LEQMLSCGER 173
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 173 IFSADKAKETPFNLAKHINSMDP-----GEETEQLKKEYVPMGVASAPLNPETA 227
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 174 GLIYVEYKRLMFRILAMRILLGCEPOLAGDSE-QQLVEAFEEETRLNLSLPIDVPSG 232
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 228 YHKLQSAATLTKTERMERKLDIKEQ-----QEEREKTEDEAMSKSDHYKRORT 282
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 233 LYRMKRAKMLHARIEINIRKIGLRASEAGQCKALQILLIHSWE-----RGRL 285
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 283 DDDLLGWLKHNSLSTEQIIDLILSLFAGHETSVAIALAIFPLQACPAVEELREEH 342
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 286 DMQ--ALKSS--STE-----LFGGHETTASATSLITLGLYPRVLOVREELK 331
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 343 EIAARAKELGESELMNDYKKMDPTOCVINEETRLGNVAFRLHRAKALDKVYKGYDIPSG 402
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 332 SKGLLCKSNQNKRLDMEILBQKYIKETRLNPNVPGGFVALKTFELNGYQIPKG 391
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 403 WKVLPVISAHLNDSRYDQPLNPNWMOQONGASSSGSSTWGNMYPGGGPRLC 462
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 392 WNVLYISICDTHDVAELFTNKEEFNPDRFMLPHRPDASRRS-----FIPGGGLRSC 442
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 463 AGSELKLEMAVFIHHLVLFKNELAE-----DDQPAFPVDP 502
| : : : : | : : : : | : : : : | : : : : | : : : : |
DQ 443 VGRKFAKVLKIFTVELARHCDQLNGPPTMKTSPIYVDNLP 487
| : : : : | : : : : | : : : : | : : : : | : : : : |

RESULT 13
YRV2_CAEEL
ID YRV2_CAEEL STANDARD; PRT; 520 AA.
AC 027514;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative cytochrome P450 CYP13A5 (EC 1.14.--).
GN CYP13A5 OR T10B9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-Bristol N2;
RA Gardner A.;

```

```

RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; 248717; CAAB8604.1; -
DR HSSP; P14779; LJPZ.
DR WormPep; T10B9.2; CE01656.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
KW BINDING 464 464 HEME (BY SIMILARITY).
FT SEQUENCE 520 AA; 59524 MW; 087B19E25B7ADF3B CRC64;
SQ
Query Match 12.1%; Score 324.5; DB 1; Length 520;
Best Local Similarity 23.6%; Pred. No. 1.7e-12;
Matches 120; Conservative 94; Mismatches 178; Indels 117; Gaps 20;
OY 42 GKSQPFILG-----ETIGYIKPYATATLGDPMQOHVSKYKITYRSLFGEPTVSAD 93
DB 33 GPRGPFELGVHIFNFDYENRGLK-----LGEWIK-----YGPITIGIEGVEKTIYSD 82
OY 94 AGLNRPILQNEGRLEECSPRSIGILG-----KWSMLVLV-GDMHRDRSISLNFSLHA 147
DB 83 P---EFVHEVFVKQDFNFGKRTNPIQGDPRNKRRAHLSAAGHMKRRLRTLSPFSSNK 139
OY 148 RLRTIL-----LKVYERITLFLVDSMOQNSIFSNODEAKKFTFNLMKHMISMDGEE 201
DB 140 NLKRTMSVTEETVELMRH---LDDMSAKGKAVDLDIYQETLIDIGRIAM---GQTE 192
OY 202 TEOLKEKYTFEMKGV-----VSAPLNLPGTAVHK-----ALQSRATILKF 241
DB 193 SLMFNPMLPKYKGIKFKOSKRLPFLVSGIFPIAGTMRERFEMRPSIQAPFLMSIVEKA 252
OY 242 IIRKAEERKLDIKE---EDQ-----EEBEVKTEDBAESKSDHYAK 279
DB 253 LNKRLRLEORADEKAGIEPSGEPQDFIDLFLDARAVNDFEESASLGFAKTEIAKYD---K 309
OY 280 QNTDDDLGWLAKHNSLSTEQILDLISLFLAGHETSSVAIALAIFLOACPKAYEELRE 339
DB 310 QLFPELILG-----OLPFLLAGYDPTALISLSSYSLARHPEIQKLOE 354
OY 340 EHLLEIARAKKEELGESELNMDYKMDFTQCVINETIRLGNVVRFLH-RKALDVAHYKGD 398
DB 355 E-----VDRECPNEVFVFDQISKLKMECVKALRYPLASLYHNKRMKMETNVLGQY 408
OY 399 IFSGKVLPLVSAVHLDNSRY-DQPLNLPWNRQOONNGASSGSGSFSTWGNNTMPREG 457
DB 409 IEKGTNVQVDWTLTHYDKRWGEDANERFRERME-----SGDELFYAKG-GYLPFGM 459
OY 458 GPRCAGSELKEMAVFIHHLVLFKNWE 486
DB 460 GPRICIGMLAMERKMLTLHLKRYTEE 488

```

```

DE Putative cytochrome P450 CYP13A7 (EC 1.14.-.-).
GN CYP13A7 OR T10B9.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Gardner A.;
RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; 248717; CAAB8609.1; -
DR HSSP; P14779; LJPZ.
DR WormPep; T10B9.10; CE01655.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
KW BINDING 464 464 HEME (BY SIMILARITY).
FT SEQUENCE 518 AA; 58999 MW; CC04283EF87B9EA7 CRC64;
SQ
Query Match 12.0%; Score 321.5; DB 1; Length 518;
Best Local Similarity 24.0%; Pred. No. 2.6e-12;
Matches 120; Conservative 96; Mismatches 184; Indels 101; Gaps 20;
OY 42 GKSQPFILG-----ETIGYIKPYATATLGDPMQOHVSKYKITYRSLFGEPTVSAD 93
DB 33 GPRGPFELGVHIFNFDYENRGLK-----FSEWIKKGPYIGIEGVEKTIYSD 82
OY 94 AGLNRPILQNEGRLEECSPRSIGILG-----KWSMLVLV-GDMHRDRSISLNFSLHA 147
DB 83 P---EFVHEVFVKQDFNFGKRTNPIQGDPRNKRRAHLSAAGHMKRRLRTLSPFSSNK 139
OY 148 RLRTI-----LKVYERITLFLVDSMOQNSIFSNODEAKKFTFNLMKHMISMDGEE 192
DB 140 NLKRTMSVTEETVELMRH---LDDMSAKGKAVDLDIYQETLIDIGRIAM---GQTE 192
OY 193 SLMFNPMLPKYKGIKFKOSKRLPFLVSGIFPIAGTMRERFEMRPSIQAPFLMSIVEKA 252
DB 188 M-----GQESLMLFRNPMYLDKVTIRKFGNNVFMISGIFRFGIALRNFAPPSIQMT 243
OY 242 IIRKAEERKLDIKEEDQ---EEBEVKTEDBAEM-----SKSDHYAKORTDDLLG 288
DB 253 LNKRLRLEORADEKAGIEPSGEPQDFIDLFLDARAVNDFEESASLGFAKTEIAKYD---K 309
OY 280 QNTDDDLGWLAKHNSLSTEQILDLISLFLAGHETSSVAIALAIFLOACPKAYEELRE 339
DB 310 QLFPELILG-----OLPFLLAGYDPTALISLSSYSLARHPEIQKLOE-----V 356
OY 340 EHLLEIARAKKEELGESELNMDYKMDFTQCVINETIRLGNVVRFLH-RKALDVAHYKGD 398
DB 355 E-----VDRECPNEVFVFDQISKLKMECVKALRYPLASLYHNKRMKMETNVLGQY 408
OY 399 IFSGKVLPLVSAVHLDNSRY-DQPLNLPWNRQOONNGASSGSGSFSTWGNNTMPREG 457
DB 409 IEKGTNVQVDWTLTHYDKRWGEDANERFRERME-----SGDELFYAKG-GYLPFGM 459
OY 458 GPRCAGSELKEMAVFIHHLVLFKNWE 486
DB 460 GPRICIGMLAMERKMLTLHLKRYTEE 488

```

DB 468 BLAMMEMKMLNLIKNTFE 488

RESULT 15
 ID C912_ARATH STANDARD; PRU: 500 AA.
 AC 065790; 09S2U3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Cytochrome P450 91A2 (EC 1.14.-.-).
 GN CYP91A2 OR AT4G37430 OR F6G17.80.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_taxid=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Seedling;
 RX MEDLINE=96281573; PubMed=9620263;
 RA Micolanti M., Ward E., Ohta D.;
 RT "Cytochrome P450 superfamily in Arabidopsis thaliana: isolation of
 RT cDNAs, differential expression, and RFLP mapping of multiple
 RL plant Mol. Biol. 37:39-52(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansoorge W., Brandt P., Grivelli L.A., Rieger M.,
 RA Weichselbacher M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis B., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Biham L., Robben J.,
 RA Van der Schueren J., Grynoprez B., Chang Y.-J., Vandenbusche F.,
 RA Braeken M., Welljens I., Voelt G., Bastlaens I., Aert R., Defoor E.,
 RA Woltzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirksen W.,
 RA Koolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Bernerster S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
 RA Petrelet A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehert T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzi A.,
 RA Neumann S., Argilou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Feilber R.,
 RA Schindl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheifons T., Weber N., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Perez-Perez A., Punelle B., Bent E., Johnson S., Tranc D., Jesse T.,
 RA Heijman L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
 RA Felsman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dechla N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Strongling T., Kallicki J., Graves T., Harmon G., Edwards J.,
 RA Latrellle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Iamir B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodi M., Johnson A.,
 RA Chen E., Maira M., Martienssen R., McCombie W.R.;

RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D76607; BAA28539.1; -
 DR EMBL: AL035601; CAB88210.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Membrane; Heme.
 FT BINDING 438 438 HEME (BY SIMILARITY).
 FT CONFLICT 106 106 A -> V (IN REF. 1).
 FT CONFLICT 127 127 I -> M (IN REF. 1).
 FT CONFLICT 140 140 N -> I (IN REF. 1).
 FT CONFLICT 454 454 S -> T (IN REF. 1).
 SQ SEQUENCE 500 AA; 57555 MW; 0FBA53D2070EA2EA CRC64;

Query Match 11.88; Score 316; DB 1; Length 500;
 Best Local Similarity 25.34; Pred. No. 5,4e-12;
 Matches 135; Conservative 83; Mismatches 199; Indels 118; Gaps 22;

QY 9 LRLPLPLSLSLFLFLILKLR--RNKRPNNPBGSGMPEIGETIGYAKPTATLGD 66
 DB 1 MLRYILLP-----LLELYISYKFLYKSTQRFNLPDPGPPSPVGH-LHMKP--PIHR 50
 QY 67 PMQOHVSKYKIRYKSNLFGPTTY-----SADAGLNRFILQNGRLFECSYPSIG 117
 DB 51 LLOYSNOYGPPIF-SLRFSGRRVYVITSPSLAQSFSGNDIYVSSPQLDTAKYAYNH 109
 QY 118 GILGKMSMLVVGDMHDMST-SLNLSHARLRT-LKQVBRHTFLVDSMOQNSIF 174
 DB 110 TVYV-----TAPYGDHNNLRICQETLSHRILNFOHKKRDELTMLTFLSYTQTS-- 163
 QY 175 SAODEAKKTF-----MIMAKHMSMDPG-----DEETPOLKKEVYTMKGV 216
 DB 164 ---NESNDFHIELEPLSLDTFNNTYRMVTKGRYYGDDVYNNKEDELFK----- 211
 QY 217 VSAPLNLPGTAYHKAQSRAT-----IKFTERKKEKRLDIKEEDDEEVEKTEDA 269
 DB 212 -----LVYDIAMVSGANHSADYLPILKLGKFKFEVKAIG-----KMDMI 253
 QY 270 EMKSDHVRKQRTDDDLGVNLKHSNLTBQILD-----LISLFRAGHETSSVAIALAI 324
 DB 254 LQALDECRDKKGNVWNLISLQOOQPYRYDVIKGLMSMMLAGETSLAVTLEMMAM 313
 QY 325 FFLQACPRAVELREHLELTARAKKEGSEL-NMDDYKMKDPTQCVINETLRLGNVAP 383
 DB 314 ANLLRNPVEYLKARSE-----IDEKIGDRDLDEEDIANVPLQNVSETFLFVYAPF 367
 QY 384 L-HRKALKVRYGYDIPSGWKVLPYISAVHLDNSRYDQNPFPWRMOQOONASSSGS 442
 DB 368 LIPRSPDMDKIGGYVPPRTIVYVAMAIHRDPEIWEEREPENPDY--NDGC----- 419
 QY 443 GSFTSGNNY-----NPFGGPRLCAGSELAKEMAVTFIHHVLYKRWELAEDDQ 492
 DB 420 -----GSDYVYKMLPFGNGRTCTGAGLIGQRTVILAGSLQCEWENVXGEE 468

Search completed: March 30, 2003, 12:05:09
 Job time : 39 secs

...

...

...

.

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2003, 11:49:16 ; Search time 63 Seconds

(without alignments)
1677.812 Million cell updates/sec

Title: US-09-502-426A-2

Sequence: 1 MFEHHHTLLPULLLPSSLIS.....FAFPVDPNGLPINVRIL 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirs:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2681	100.0	513	10	064989	064989 arabidopsis
2	2677	99.9	513	10	09SC09	09SC09 arabidopsis
3	951	35.5	474	10	09LKH7	09LKH7 vigna radia
4	866	32.3	512	10	09FX29	09FX29 arabidopsis
5	843.5	31.5	491	10	0941A6	0941A6 arabidopsis
6	822	30.7	478	10	09LNT3	09LNT3 arabidopsis
7	815	30.4	382	10	09LX89	09LX89 arabidopsis
8	813.5	30.3	490	10	0941W5	0941W5 oryza sativ
9	789	29.4	464	10	09LIC5	09LIC5 arabidopsis
10	756.5	28.2	481	10	08W0R4	08W0R4 sorghum bic
11	726	27.1	465	10	09FMA5	09FMA5 arabidopsis
12	712	26.6	465	10	0940V4	0940V4 arabidopsis
13	711	26.5	465	10	09LH81	09LH81 arabidopsis
14	672.5	25.1	467	10	0949P1	0949P1 arabidopsis
15	658	24.5	463	10	09PH76	09PH76 arabidopsis
16	632	23.6	457	10	065624	065624 arabidopsis

17	631	23.5	518	10	09FI38	09FI38 arabidopsis
18	631	23.5	735	10	09LGI7	09LGI7 arabidopsis
19	622	23.2	443	10	091JK2	091JK2 arabidopsis
20	611	22.8	482	10	081077	081077 arabidopsis
21	606	22.6	479	10	08VZC2	08VZC2 arabidopsis
22	597.5	22.3	485	10	08W4T9	08W4T9 arabidopsis
23	596.5	22.2	497	10	09AXM6	09AXM6 taxus cuspi
24	589	22.0	486	10	09LIT69	09LIT69 arabidopsis
25	581	21.7	487	10	023384	023384 arabidopsis
26	581	21.7	496	10	09FOY4	09FOY4 cucurbita m
27	579.5	21.6	474	10	09SHY7	09SHY7 arabidopsis
28	579.5	21.6	485	10	09SHR2	09SHR2 arabidopsis
29	577	21.5	455	10	09LKH8	09LKH8 arabidopsis
30	576.5	21.5	490	10	09C5Y3	09C5Y3 arabidopsis
31	575.5	21.5	464	10	004949	004949 arabidopsis
32	569.5	21.2	477	10	09LVT7	09LVT7 arabidopsis
33	547.5	20.4	489	10	09ZV72	09ZV72 arabidopsis
34	546.5	20.4	444	10	08S917	08S917 arabidopsis
35	546.5	20.4	489	10	09C5Y2	09C5Y2 arabidopsis
36	518	19.3	504	10	09SNG3	09SNG3 oryza sativ
37	502	18.7	499	10	09AXH9	09AXH9 hordeum vul
38	497.5	18.6	460	10	09SYN2	09SYN2 arabidopsis
39	453	16.9	497	11	08VIT0	08VIT0 rattus norv
40	435	16.2	224	10	09SDM6	09SDM6 helianthus
41	421	15.7	394	10	08WIS6	08WIS6 zea mays su
42	414	15.4	394	10	08WIT1	08WIT1 zea mays su
43	414	15.4	394	10	08WIS9	08WIS9 zea mays su
44	413	15.4	394	10	08WVJ5	08WVJ5 zea mays su
45	412	15.4	394	10	08WIV5	08WIV5 zea mays su

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	513 AA.
064989			
ID 064989			
AC 064989;			
DT 01-AUG-1998 (TREMBLrel. 07, Created)			
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)			
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)			
DE Steroid 22-alpha-hydroxylase.			
GN DWF4.			
OS Arabidopsis thaliana (Mouse-ear cress).			
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC eurooids II; Brassicales; Brassicaceae; Arabidopsis.			
OX NCBI_TaxID=3702;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=MS-2;			
RX MEDLINE=98158690; PubMed=9490746;			
RA Choe S., Dikes B.P., Fujioka S., Takatsuto S., Sakurai A.,			
RA Feldmann K.A.;			
RT "The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates			
RT multiple 22alpha-hydroxylation steps in brassinosteroid			
RT biosynthesis."			
RT Plant Cell 10:231-243(1998).			
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
DR EMBL: AF044216; AAC05093.1; -;			
DR InterPro: IPR001128; Cytochrome_P450.			
DR Pfam: PF00067; P450; 1.			
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.			
KW Heme; Monooxygenase; Oxidoreductase.			
SO SEQUENCE 513 AA; 58867 MW; B1639BDD9A4DA6F3 CRC64;			
QY Query Match 100.0%; Score 2681; DB 10; Length 513;			
Db Best Local Similarity 100.0%; Pred. No. 1.6e-184;			
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
1 MFEHHHTLLPULLLPSSLISLFLILKRRNRKTRNLPFGSGWFFLGCTIGLYKPYT 60			
1 MFEHHHTLLPULLLPSSLISLFLILKRRNRKTRNLPFGSGWFFLGCTIGLYKPYT 60			

```

QY 61 ATTLDGFMQOHXSKYKGYRSMLEGEPTIVSADAGLNRIILONEGRLEFCSTPRISGTL 120
DB 61 ATTLDGFMQOHXSKYKGYRSMLEGEPTIVSADAGLNRIILONEGRLEFCSTPRISGTL 120
QY 121 GKMSMLVYVGDMHMRMSISLNFSLHARLRTLLKDVREHTLFLVDSMOQNSIFSAODEA 180
DB 121 GKMSMLVYVGDMHMRMSISLNFSLHARLRTLLKDVREHTLFLVDSMOQNSIFSAODEA 180
QY 121 GKMSMLVYVGDMHMRMSISLNFSLHARLRTLLKDVREHTLFLVDSMOQNSIFSAODEA 180
DB 121 GKMSMLVYVGDMHMRMSISLNFSLHARLRTLLKDVREHTLFLVDSMOQNSIFSAODEA 180
QY 181 KKFETNLMAKHIMSMDPGSEETEOLKKEYVPMKVSAPLMPTAYHKAQSATILK 240
DB 181 KKFETNLMAKHIMSMDPGSEETEOLKKEYVPMKVSAPLMPTAYHKAQSATILK 240
QY 241 FIERMEERKLDIKKEDEEVEEKTEDAEKMSKSHVRKQRTDDLLGVLKHSNLSTEQ 300
DB 241 FIERMEERKLDIKKEDEEVEEKTEDAEKMSKSHVRKQRTDDLLGVLKHSNLSTEQ 300
QY 301 ILDLISLFLFAGHETSSVAIALAIFFLQACPKAVELEBEHEIARAKKEGESELMND 360
DB 301 ILDLISLFLFAGHETSSVAIALAIFFLQACPKAVELEBEHEIARAKKEGESELMND 360
QY 361 YKKMDFTCVINEITRLGNVVFELHRRKALKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
DB 361 YKKMDFTCVINEITRLGNVVFELHRRKALKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
QY 421 QPNLEPNRMWQOQNGASSSGSSTGWNNTMPGGGPRCLAGSELAKLEMAVFIHHLY 480
DB 421 QPNLEPNRMWQOQNGASSSGSSTGWNNTMPGGGPRCLAGSELAKLEMAVFIHHLY 480
QY 481 LKFNNELAEDDQPFAPFPVDFPNGPLPIRVSRIL 513
DB 481 LKFNNELAEDDQPFAPFPVDFPNGPLPIRVSRIL 513

```

RESULT 2

```

Q9SCQ9 PRELIMINARY; PRT; 513 AA.
AC Q9SCQ9:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Steroid 22-alpha-hydroxylase (DMP4) (A13950660/T3A5_40).
GN T3A5_40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RA Bloecher H., Mewes H.W., Lemcke K., Mayer K.F.X., Quettier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamlya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN 14
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shin P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,

```

```

RA Hayashizaki Y., Ishida J., Jones T., Kamlya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AL132979; CAB62435.1; -.
DR EMBL: AF412114; AA106567.1; -.
DR EMBL: AY090266; AA190927.1; -.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450.1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KM Heme: Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58667 MW; B1639DD9A5D7C93 CRC64;

```

Query Match 99.9%; Score 2677; DB 10; Length 513;
Best Local Similarity 99.8%; Pred. No. 3e-184;
Matches 512; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MEFEHHTLLPLLLPSLISLFLFLILKRRNRKTRPNLPPEKSGWPLGETIGYLPYT 60
DB 1 MEFEHHTLLPLLLPSLISLFLFLILKRRNRKTRPNLPPEKSGWPLGETIGYLPYT 60
QY 61 ATTLDGFMQOHXSKYKGYRSMLEGEPTIVSADAGLNRIILONEGRLEFCSTPRISGTL 120
DB 61 ATTLDGFMQOHXSKYKGYRSMLEGEPTIVSADAGLNRIILONEGRLEFCSTPRISGTL 120
QY 121 GKMSMLVYVGDMHMRMSISLNFSLHARLRTLLKDVREHTLFLVDSMOQNSIFSAODEA 180
DB 121 GKMSMLVYVGDMHMRMSISLNFSLHARLRTLLKDVREHTLFLVDSMOQNSIFSAODEA 180
QY 181 KKFETNLMAKHIMSMDPGSEETEOLKKEYVPMKVSAPLMPTAYHKAQSATILK 240
DB 181 KKFETNLMAKHIMSMDPGSEETEOLKKEYVPMKVSAPLMPTAYHKAQSATILK 240
QY 241 FIERMEERKLDIKKEDEEVEEKTEDAEKMSKSHVRKQRTDDLLGVLKHSNLSTEQ 300
DB 241 FIERMEERKLDIKKEDEEVEEKTEDAEKMSKSHVRKQRTDDLLGVLKHSNLSTEQ 300
QY 301 ILDLISLFLFAGHETSSVAIALAIFFLQACPKAVELEBEHEIARAKKEGESELMND 360
DB 301 ILDLISLFLFAGHETSSVAIALAIFFLQACPKAVELEBEHEIARAKKEGESELMND 360
QY 361 YKKMDFTCVINEITRLGNVVFELHRRKALKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
DB 361 YKKMDFTCVINEITRLGNVVFELHRRKALKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
QY 421 QPNLEPNRMWQOQNGASSSGSSTGWNNTMPGGGPRCLAGSELAKLEMAVFIHHLY 480
DB 421 QPNLEPNRMWQOQNGASSSGSSTGWNNTMPGGGPRCLAGSELAKLEMAVFIHHLY 480
QY 481 LKFNNELAEDDQPFAPFPVDFPNGPLPIRVSRIL 513
DB 481 LKFNNELAEDDQPFAPFPVDFPNGPLPIRVSRIL 513

```

RESULT 3

```

Q9LKH7 PRELIMINARY; PRT; 474 AA.
AC Q9LKH7:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cytochrome P450.
GN CYP90A2.
OS Vigna radiata.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.

```

OX NCBI_TaxID=157791;
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV_2937;
 RA Yang M.T., Chen Y.M.;
 RT "Cloning and sequencing of a *Vigna radiata* cDNA encoding cytochrome
 P450."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: A27252; AAF89209.1;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 474 AA; 54037 MW; A2542A809C5BAC6D CRC64;

Query Match 35.5%; Score 951; DB 10; Length 474;
 Best Local Similarity 40.1%; Pred. No. 3e-60;
 Matches 203; Conservative 97; Mismatches 162; Indels 44; Gaps 10;

QY 12 LLLPSSLLELLF-----LILKRRNRKTRFNLPPGKSGWPLIGETIGYLPYATATLGD 66
 Db 1 MWSLPTL-LLLFASAAAFILHAFSRKFRPLPGSYGLPFIETGLISAYKSSNPP 58
 QY 67 FMOQHVKYKTYRSNLFGEPTIVSADAGLNRTIIQNGERLEFECSTPSIGILKWSKL 126
 Db 59 FVDEVRKRGSTFMTNHFGEPTVFSADPELNRFILQNEGLKLDGSPGISMLKHSLL 118
 QY 127 VLVDGNDHMRISINFLSHARLRTLLKDYERHTLFVDSMQNSIFSADAEKKTEN 186
 Db 119 LKGAHLKRMHSLTMSFANSIITKHLLHIDRLGLNLDTSR--VTLMDQAKITFE 176
 QY 187 LMAKHMSMDPEETEOALKRYTFMGQVSAPLNLPGLAHKALQSRATILKTERKM 246
 Db 177 LTVKQMSMDP-DEWTESLRKYLYIEGFTLLPLFSTYRAIKATVAAELTLV 235
 QY 247 EERKIDKEEDQEEVEVTEDEAEKSDHVRKQRTDIDLGLWYL-KISNLTEDILDI 305
 Db 236 RQR---REEVNQGKEKS-----DMGLALASGDHFSDDIYVFL 272
 QY 306 LSLFAGHETSSVAIALAIFLQACPKAVEELREEHLEIARAKKELGSELMNDYKMD 365
 Db 273 LALLVAGETSTTMTLVKFLTEPLALQLEKHDI-RAKSDPG-APLEWTDYKMW 330
 QY 366 FTQCYNFETLRGNVRFHLRKALKDYKGYDIPSGKVLPIVSAVHLNDSRYDQPLF 425
 Db 331 FTQHVNETLVANIGIFERRATTDIDIKGYTTPKWKYASFRVAILNDEYKDAATF 390
 QY 426 NPMWQOQNGASSSGSFSSTGNNYVFGGPRLCAGSELAKLEMAVFTHLVLFKNW 485
 Db 391 NPMWQSNSSBAAP-----ANYVTPFGGPRLCGYELAVLVSVFLHRTVTRFSW 442
 QY 486 ELADDQEPAPFPYDFPENGPIRYSR 511
 Db 443 VPAEDKLVFPFTTRTKRYPPIYKR 468

RESULT 4
 Q9FX29 PRELIMINARY; PRT; 512 AA.
 AC Q9FX29;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
 DE Steroid 22-alpha-hydroxylase, putative.
 GN T9L24.44.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=CV_COLDIMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
 RA Maiti R., Romling C.M., Koo H., Fujii C.Y., Utterback T.R.,
 RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome 1 BAC T9L24 genomic sequence."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AC012396; AAG30983.1;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 512 AA; 58297 MW; 1590C0304BF319PA CRC64;

Query Match 32.3%; Score 866; DB 10; Length 512;
 Best Local Similarity 33.8%; Pred. No. 4.2e-54;
 Matches 181; Conservative 105; Mismatches 173; Indels 76; Gaps 7;

QY 12 LLLPSSLLELLF-----LILKRRNRKTRFNLPPGKSGWPLIGETIGYLPYATATLGD 66
 Db 11 LVSSTTFIAFIIFLLAGIARKRRAHRLPPGSGWPLIGDTFAMLVAVAGSHSS 70
 QY 67 FMOQHVKYKTYRSNLFGEPTIVSADAGLNRTIIQNGERLEFECSTPSIGILKWSKL 101
 Db 71 FVEQIKFVSLGSLVLLILKRPDNGSFNIRYGRIFSCGLFKMAVVSADPPFNRI 130
 QY 102 QNEGRLEFECSTPSISIGILKWSMLVLVDGNDHMRISINFLSHARLRTLLKDYERHT 161
 Db 131 QNEGRLEFECSTPSISIGILKWSMLVLVDGNDHMRISINFLSHARLRTLLKDYERHT 190
 QY 162 LFVDSMQNSIFSADAEKKTENLMAKHMSMDPEETEOALKRYTFMGQVSAPL 221
 Db 191 LQTLNFKRGVYVLLQDCRKVAIHLWNLQVLS-SESEVDEMSQLFSPFVQGLSVPI 249
 QY 222 NLPSTAVHKAQSRATILKTERMEERKIDKEEDQEEVEVTEDEAEKSDHVRKQR 281
 Db 250 DLPGFYTNKAKARKETIRKINKTERLQNKASD-----T 286
 QY 282 TDDDLGWLKHSNLSTEOILDLISLFPCHETSSVAIALAIFLQACPKAVEELREEH 341
 Db 287 AGNSVLRLEELSELPESNADEFLINLIFAGNETSTKMLFAVYFLTHCRKAMTQLEEH 346
 QY 342 LETARAKKEGESELMNDYDKKMDFTQCYNFETLRGNVRFHLRKALKDYKGYDIP 401
 Db 347 -----DRLAGGLTMDQYKTDFTQCYIDETLRGLGAIWLMREKKEVSYQDYVIRK 399
 QY 402 GMYLPIVSAVHLNDSRYDQPLFNPWR-----QOQNGASSSGSFSSTGNNYMPG 456
 Db 400 GCFVVPFLSAVHLDESYKESLSEFNPRWMLDPETQCKRMNRISP-----FTCPFG 449
 QY 457 GGPRLCAGSELAKLEMAVFTHLVLFKNWELADDQEPAPFPYDFPENGPIRYSR 511
 Db 450 GGRFCPGAEALAKQIALFLHFYITTYKWTQLEKEDISFPSPARLVNGFKIQINR 504

RESULT 5
 Q94IAG PRELIMINARY; PRT; 491 AA.
 AC Q94IAG;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE CYP90D.
 GN CYP90D.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.

RC	STRAIN=CV, COLUMBIA;
RA	Shimada Y.;
RT	"P450 gene repressed by brassinosteroid.";
RL	submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
CC	-1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR	EMBL; AB066286; BAB62109.1; "
DR	InterPro: IPR001128; Cytochrome P450.
DR	InterPro: IPR000504; RNA_rec_mot.
DR	Pfam; PF00067; P450; 1.
DR	PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN.1.
DR	PROSITE; PS00030; RM_RNP_1; UNKNOWN.1.
KW	Heme; Monooxygenase; Oxidoreductase.
SO	SEQUENCE 491 AA; 56153 MW; 02FEB98A91995AA0 CRC64;
Query Match	31.5%; Score 843.5; DB 10; Length 491;
Best Local Similarity	37.0%; Pred. No. 1.7e-52;
Matches 176; Conservative 104; Mismatches 157; Indels 39; Gaps 9;	
QY	40 PPGKSGMPFLGTCIGYLPRYATTLGDMQOVSVYGYKRNRLGEPYVSADGLNRF 99
DB	52 PHGSLGMPVIGETLIFEVSSANS DREPSTMDKRLRMIGRFKSHITGATVYSTDEVNNA 111
QY	100 ILONGRLFECSYPSRISGILGKMSLVYGDMDHDMRSISINTFLSHRLRTLLKDYER 159
DB	112 VLQDSSTAFVPPYPTVAVELMGKSSILLNGLSHRFGVLSFLSKPLKQIARDHK 171
QY	160 HTLEVLDSWQONSIFSAODEAKKFTFNMAKHIMSMDPEEETEDLKREYTFMGVSA 219
DB	172 FLSESMDSMEQOPLLDVDSKTVAFKYALKALISVEKG-EDLEELKREFENFISGLMSL 230
QY	220 PLNLGCTAYHNRKLGSRATILKFERKMERKLDIKEEOEEVEYTPEDAEV---SKSPH 276
DB	231 PINFGTOLHNSLQKKMKNVCOVERIIEKIR-RTKNKEEDVIAKDVVDVLKDSSEH 288
QY	277 VRKQRTDDLLGWLKHSNSTEQTLIDLTLISLFFAGHETSVAIALAIFLOACEKAAVE 336
DB	289 -----LTH-NLIANNMDMMI-----PGHDSVPLITLAKVFLSDSPAALNL 329
QY	337 LREEHLEIARAKKELGESEBLNMDYDKKADFTQCVINETRLGAVNVRFLHRAKDKVRYKG 396
DB	330 LTEENMKL-KSLKELTGEPILWYNDLISLPFOKVTITETLRMGVNIIGVRRKMKDVEIKG 388
QY	397 YDIPSGKVLVPLVPSAVHLDNSKYDDPNLFNPRMOQONNGASSSGSSTGMNYPMPG 456
DB	389 YVIRPGWCFLLVNLNSVHLIDKLKLYESPYFNFMKROERDMNTSS-----FSPG 436
QY	457 GGPRLCAGSELAKEMAVFIHHLVLFKFWMLAEADQPAFPVDPENGLPIRVSRH 512
DB	437 GGORICPELDLARLETSVFLHILVTRFRW-IAEDTTINFPVHMKKPLPIRKRI 491

RT	FT	1	Genomic sequence for Arabidopsis thaliana BAC T1C24 from chromosome
RT	1	I	"
RL	Submitted (JAN-2001)		to the EMBL/GenBank/DBJ databases.
CC	-1	STABILITY:	BELONGS TO THE CYTOCHROME P450 FAMILY.
DR	EMBL; AC025417; AAF8087.1;		-
DR	InterPro: IPR001128; Cytochrome_P450.		
DR	Pfam: PF00067; p450; 2.		
DR	PRINTS: PR00385; p450.		
DR	PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.		
KW	Heme; Monooxygenase; Oxidoreductase.		
KW	SEQUENCE 478 AA; 55058 MW; F67A999F2FD28558 CRC64;		
SO	Query Match	30.7%;	Score 82; DB 10; length 478;
	Best Local Similarity	35.6%;	Pred. No. 5,6e-51;
	Matches 181; Conservative	95;	Mismatches 187; Indels 46; Gaps
QY	12	LLLLPSLLSLLFLILKRRNRKTRFNPRLPGKSGMPFLGETIGLYKPYATITLGDPMOQH	71
DB	4	LLIIVSLIISTTHWVSWNRNPKCGKGLPPSGMGPELVLSIQFRRPNSTDISPEIKRK	63
QY	72	VSKYGIKYSNLFGEPTIYASADAGNRFILQEGRLSECSYPSRISGIGIKKSMULVYD	131
DB	64	VKKYGIPIKTNLVGRPVIVSTADLSYFENDEGRCFOSWYDTFTHIGKKNVGSLSHF	123
QY	132	MHRDYSISLNFLSHARLRTILKDYERTTFLVDSMOQNSIFSADDEAKKTFNIMAKH	191
DB	124	MYKYLKNWVLTGFGDGKTK-MLPQVEMTANKRLIEIWSNODSVELKDATAISMIFDLTAK	182
QY	192	IMSNDGGEETPOLKEYTEFMKGVSAPLNPGRVYHNLASDRAITLFIKRMERKL	251
DB	183	LISHDP-DKSSNLANVAVFTQGLISPEFDLPGRVYHNLASDRAITLFIKRMERKL	240
QY	252	DIKEEDQEEVEKYTEDEDEMSKSDHYRKQRTD-DLLGLVNLKNSLTQEI-LDLISL	308
DB	241	-----ENPRKNPSPDFVDVIEIQEGITLIEIALDLMLFL	277
QY	309	LEAGHTSSVALALIFLQACPKAVBELREHNLIAKKKELGSELNWDYDKKDFQ	368
DB	278	LFASSETISLALTIAIKPLSDPEVLKRLTEHEHTLR-NRSDAOSGLWEYKSKSYTF	336
QY	369	CVINETLRLGNVRLRHRAKLDVRYK-----GYDIPSGMWVLPIVASVHNDNSRYDOP	422
DB	337	QFINETARLANIVPAIFKRALDIFKEKEVNDTDTIPGMVAVMCPRAVHLNPEYKRP	396
QY	423	NLFNWRMQOQNGASSGSGSFSTWGNANNYPFGGPRICAGSELAKLEMAVFIHHLVAK	482
DB	397	LVFNISRME-----GSKVTNASKRFMAFGGMRCVGTIDTKLQMAAFHLSHLVYK	446
QY	483	FNWELAEDDQPFAPFVDFPNGLPIRVSR	511
DB	447	YRMEIKGGINTRTPGLQFPNGYHVKLHK	475
RESULT 7			
ID	Q9LY89	PRELIMINARY;	PRT; 382 AA.
AC	Q9LY89;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Hypothetical 43.9 kDa protein.		
GN	F18022.190.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsids.		
NCBI	taxid=3702;		
LN	11		
RP	SEQUENCE FROM N.A.		
RA	Beyan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,		
RA	Rudd S., Demko K., Mayer K.F.X.;		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
LN	[2]		

RP SEQUENCE FROM N.A.
 EU Arabidopsis sequencing project.
 Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AL163817; CAB87779.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; p450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme: Hypothetical protein; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 382 AA; 43869 MW; 1BC5685AB734BE30 CRC64;

Query Match 30.4%; Score 815; DB 10; Length 382;
 Best Local Similarity 34.2%; Pred. No. 1.3e-50;
 Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

QY 45 GMPFGETIGYKPTATTLGDPQOQHVSKGKTYRSLNFGEPITVSADAGINREITLONE 104
 DB 2 GMPFGETISFPKPHRSISIGFTLQORVSRGKPKNSGCGKAVVSCQDELNMFILONE 61
 QY 105 GRLFCSPRSISGILGKMSLVYGDHMRRSISLNFSLHARLRLTLKDYENHTLFV 164
 DB 62 GRLFTSDPKAMHDILGKYSLLATGELHRLKLVITISLITKSKPDLHCAENLSISI 121
 QY 165 LDMQONSIFSAQDEAKKFTFNLAHMSMDPGEETEOLEKKEVYTMKGVSAPLNP 224
 DB 122 LKSMKCEVEERHKEVKFTLSVMVNOQLSTIKPEDPARLYLQDPLSMKGISLPIPLP 181
 QY 225 GAAYKALQSRATILKFERKMEERKLDIKEDDEEVEKTEDEAEMSKSDHVRKQRTDD 284
 DB 182 GTGYNALAKVSNRNHONATIEDNNNAIREDFLDSITSNED----- 224
 QY 285 DLGAVLNKSNLSTEOIIDLILSLFAGHETSSVAIALAIFLQACRAVELREHLEI 344
 DB 225 -----EEHRAI 230
 QY 345 ARAKKEGESELMNDYKAKMDFQCYINETLRGNVAFELHRAKLDVRYKGYDIPSGWK 404
 DB 231 -RAKKGDEL-LNMEEDYOKMEFTOCVISEALRCGNIVYTVARAKHDIKNEYIYPRGKK 288
 QY 405 VLPVISAHLNDRYDQNLFPNWRMOQNNNGASSSGSFTMGNNMPPRGCGPRICAG 464
 DB 289 VFPTFVNHDLPSLHNPFEENPRKWTYT-----AGGGVAVRCGP 329
 QY 465 SELAKEMAVFIHHLVLEKFMELAEDDQPAFPVDPNGIPRIV 509
 DB 330 GELCKLQIAFPLHLVLSYRWKIKSDEMPLAHNPVEFKRGLLHI 374

RESULT 8

ID 0941W5 PRELIMINARY; PRT; 490 AA.
 AC 0941W5;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Cytochrome P450-like protein.
 GN P0419B01.5.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriharoidae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone: P0419B01.
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: A0003244; BAB56089.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; p450; 1.

DR PROSITE: PS00086; CYTOCHROME P450; UNKNOWN_1.
 KW Heme: Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 490 AA; 54824 MW; 9EC2853BBAFB88F CRC64;

Query Match 30.3%; Score 813.5; DB 10; Length 490;
 Best Local Similarity 34.2%; Pred. No. 2.4e-50;
 Matches 173; Conservative 106; Mismatches 156; Indels 71; Gaps 9;

QY 27 LKRRNR-----KRFNLPKSGMPLGFTIGYKPTATTLGDPQOQHVSKYK 77
 DB 28 LKRRRRAGSGKGDAAAARLPKPSFGVPVAGTELEFVSCAYSPPRAFPDKRRKLHGS 87
 QY 78 -LYRNLNFGEPITVSADAGINRFLQNEGRLEFCSPRSISGILGKMSLVYGDHMRDM 136
 DB 88 AVFRRHLSGANTVYADAEVRFVLSADARAFVWYTPRSLSLELKGSLILNALQRRV 147
 QY 137 KSIISNLSHARLTLKDYENHTLFVLDWMQONSIFSAQDEAKKFTFNLAHMSMD 196
 DB 148 HGLVGAFFKSSHLKSQLDADMRRLSPALSSFPDSSLHVGHLAKSVFEILVGLGLE 207
 QY 197 PGEETEOLEKKEVYTMKGVSAPLNPJGTAYHRAQSRATILKFERKMEERK----- 250
 DB 208 AG-EBMQOLKQFOEFYIGLMSLPKLPGLRYLSLQAKKMAHLQRIIREKARRAA 266
 QY 251 -----LDIKEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGAVLNKSNLSTEOIIDL 304
 DB 267 SPPRADIVLIGDSDE-----LTDELISDN 292
 QY 305 ILSLFAGHETSSVAIALAIFLQACRAVELREHLEIARAKKEGESELMNDYKMK 364
 DB 293 MIDMLPAEDSVPLITLAVKFLSECPALHQLEENIQLRKRTDMGET-LQWTQVMSL 351
 QY 365 DFTOCVINETLRGNVAFELHRAKLDVRYKGYDIPSGMYLPVISAHLNDRYDQNL 424
 DB 352 SFTGHVITETLRGNITIGIMKRAVRYEYKGHILPCKMCFYFRVSHLDLTYDEPYK 411
 QY 425 FNPWRMOQNNNGASSSGSFTMGNNMPPRGCGPRICAGSELAKEMAVFIHHLVLEKFN 484
 DB 412 FNPWRMEKED-----MSGSFT-----PFGGQRLCPGLDARLEASIFLHLVTSFR 459
 QY 485 WELAEDDQPAFPVDPNGIPRIVS 510
 DB 460 W-VAEEDHYNFPVRLKRPPIRYT 484

RESULT 9

ID 091IC5 PRELIMINARY; PRT; 464 AA.
 AC 091IC5;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Cytochrome P450-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the regions of 4,251,695 bp covered by ninety pl, TAC and BAC clones."
 RT DNA Res. 7:217-221(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL: AP001307; BAB01922.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00067; P450.1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE: PS00030; RNA_RNP_1; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 464 AA; 52864 MW; A65E094665E5190B CRC64;

Query Match 29.4%; Score 789; DB 10; Length 464;
 Best Local Similarity 36.5%; Pred. No. 1,3e-48;
 Matches 163; Conservative 101; Mismatches 145; Indels 38; Gaps 8;

QY 40 PRKSGKGPPLGTTGTLKPYATTTGDEMOQVSKYKTYRNSLFGEPITVSADAGLNF 99
 DB 52 PHSGLGWPVIGETIEFVSAYSDRESEFMDKRRRLMYGRVFKSHIGTAVTSDAEVNA 111
 QY 100 ILONGRLFECSYPSISGILGKMSMLVYGDHMDMSISLNFSLHARLRTILKDYER 159
 DB 112 VLSQSTAVFPPTVRELMGKSSILTLNGSLHRRFGLVGSFLKSPLLKAQIVADMK 171
 QY 160 HTLFVDSMOQNSISADDEAKFTFNLMKAIMSDPGEETEQDLKREYTFMKGVSA 219
 DB 172 FLSEMDLMSDQPVLLDQVSKTAVFKYLAALISVEKG-EDLELKRFEFVIGLSML 230
 QY 220 PLNPGTAVHKAQSRATLTKFERKMEERKIDKEEDQEEVEKTEDEAEN---SKSDH 276
 DB 231 PINFGTQHLRSRQAKKNKQVERIIEBKIR--KTKNKEEDVIAKLDVAVDLKDSFH 288
 QY 277 VAKQTDDDLGLVWLKHSNLSTEQDLDLILSLFAGHESVAIALAFELQACRYAE 336
 DB 289 -----LTH-NLIANNMIDMI---PGHDSVPLTLTAKVFLSDSPAALNL 329
 QY 337 LREHELEIARAKKEGESELNMDYKKMDFTQCVINETLGLGVNRFELRKALDVRXK 396
 DB 330 LEEEMAKL-KSLKELTGEELVNDYLSLPTQKVTETIRKQNVITIGVRRKMKVEIKG 388
 QY 337 YDIEGKVLVYISAVHLDNSRYDQNLFPWRMOQONNGASSSGSFSWGNMYPFG 456
 DB 389 YVIFGWFGLATLRSVHLDKLYESPYKFNPMQERDMNTSS-----FSDFG 436
 QY 457 GGPRLCAGSELAKLEMAVFHHVLKLF 483
 DB 437 GGQRLCPGLDLAKLETSVFLHHLVTRF 463

RESULT 10
 O9MORA PRELIMINARY: PRT: 481 AA.

AC O9MORA.
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative cytochrome P450-like protein.
 GN SB32H17.4.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Sorghum.
 OX NCBI_TaxID=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. BTK623;
 RA Park Y.-J., Ramakrishna W., Sanmiguell P., Emberton J., Bennetzen J.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. BTK623;
 RA Llaa V., Young S., Kovchok S., Messing J.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF466201; AAL73972.1; -
 DR InterPro: IPR001128; Cytochrome_P450.

PK Pfam: PF00067; P450.1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 SQ SEQUENCE 481 AA; 54984 MW; 6145CEPFOFD4FE CRC64;

Query Match 28.2%; Score 756.5; DB 10; Length 481;
 Best Local Similarity 33.5%; Pred. No. 2.9e-46;
 Matches 170; Conservative 104; Mismatches 190; Indels 43; Gaps 7;

QY 15 LPSLLSLFLILK-----RRNRKTRNLPKSGKMPLEGTIGYLPYATTTG 65
 DB 6 LAALSVTLGAILLHMAFKMNYGRGEGMLPPSGSLPELDETLEFFAASPTLELV 65
 QY 66 DMQOHVSKYKTYRNSLFGEPITVSADAGLNRFLQNGRLFECSYPSISGILGKSM 125
 DB 66 PEFKRLERFEGPIFRNTIVGEDMIVSLDELNARVLOOEENGFQIWTYSSFRITIGADM 125
 QY 126 ILVYGDHMDMSISLNFSLHARLRTILKDYERHTLFVDSMOQNSISFSADEAKKFTF 185
 DB 126 VSMGLPRLHRIINLVRLFGPEALNLVLLRDYORSARDELRSWLDREVEYVTTATSRMTF 185
 QY 186 NMAKHMSPDGEETEQDLKREYTFMKGVSAPLNPGTAVHKAQSRATLTKFERK 245
 DB 186 GVTAKKLISHD-DVAGSGSLMKCFDAMTKGLSPICVGTAFYRCMGGRKVMKVLKQ 244
 QY 246 MEERLIDKEEDQEEVEKTEDEAKSKSDHYRKQRTDDDLGLVWLKHSNLSTEQI-LDL 304
 DB 245 LDERR-----NGAERTYDFDL-----VIDELDKPMSISEIALNL 282
 QY 305 ILSLFAGHESVAIALAFELQACRYAEELREHELEIARAKKEGESELNMDYKKM 364
 DB 283 LFLFLFASHETTSKGLTYLKLTLNPKSLQDLTEHEHMERRDV-PSDITWEYSYM 341
 QY 365 DFTQCVINETLGLGVNRFELRKALDVRXKGYDIPSGKVLVYISAVHLDNSRYDQNL 424
 DB 342 KTSVHIESRLANLAPVRFQADQVHIKQTYLPEBSKIMICSAHLNSKYVEDPLA 401
 QY 425 FNPWRMOQONNGASSSGSFSWGNMYPFGGGRRLCAGSELAKLEMAVFHHVLKFN 484
 DB 402 FNPWRMKTPEFVGS-----KDFMAFGGLRCLVGAEFAKLQAMFLHYLVTRF 452
 QY 485 WELAEDDQFAPFPVDFPGLPIRYSR 511
 DB 453 WKALSKGTMLYPLGLRFPDGFHIQLHK 479

RESULT 11
 O9FMAS PRELIMINARY: PRT: 465 AA.

AC O9FMAS.
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Cytochrome P450 (Brassinosteroid-6-oxidase).
 GN BROOX1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=98290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Physically assigned pl and TAC clones";
 RL DNA Res. 5:41-54(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=21295570; PubMed=11402205;

RA Shinada Y., Fujioka S., Miyauchi N., Kushihiro M., Takatsuto S.,
 RA Nomura T., Yokota T., Kamiya Y., Bishop G., Yoshida S.,
 RT "Brassinosteroid-6-oxidases from Arabidopsis and tomato catalyze
 multiple C-6 oxidations in brassinosteroid biosynthesis.",
 RL Plant Physiol. 126:770-779(2001).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AB009048; BAB08653.1; -
 DR EMBL: AB035868; BAB08658.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 465 AA; 53767 MW; 0C00459C0866D1F CRC64;
 Query Match 27.1%; Score 726; DB 10; Length 465;
 Best Local Similarity 32.3%; Pred. No. 4.3e-44;
 Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;
 QY 12 LLLPSTLSLFLILKRRN--KTRRFLPGKSGWPLGTTIGYLRPYATTGDMQ 69
 DB 11 LLLPSTLSLFLILKRRN--KTRRFLPGKSGWPLGTTIGYLRPYATTGDMQ 69
 QY 70 QHVSXKGIYRSNLEGEPTVSADAGLNRFILNNGRLFECSYPSISGILGKSMVLV 129
 DB 60 NQRLRGSEFKSHLIGCPILISMDESVNRKILAKNSKGLVPGYPOSMDLIGTCMAAVH 119
 QY 130 GMDHDMSSISLNTSHARLRTILKDYERHTLFYLDSSMOONSISADDEAKKFTF 185
 DB 120 GSHRLMKSLSLISSTMMRDHILPKYDHEFRSYLDQMLEVLDIDDKTHMAFLSL 179
 QY 186 -----NMAKHIMSDPGEETEOLEKKEVYTFMGVAPLNPCTAYHKAQSRATILK 240
 DB 180 TQIANLKKRPV-----EETKAFKLVGTLISVPLIDLPCTNRCGIQANNND 229
 QY 241 FIERMEERKLDKEEDOEKEEVKTEDEAEMSKSDHVRKORTDDLLGWVLKHSN--L 296
 DB 230 LIRELMQERR-----DSGE-----TFDDMLGYLKKKEGNRP 262
 QY 297 STEOILDLISLFGHETSSVAIALAIFFLQACRAVEERREHLEIARAKKEGSEL 356
 DB 263 TDELRDQVTLISGYETVSTSMALIKYLHDKALQELRAELAFERERKQ--DEPL 320
 QY 357 NMDYKKMDFOCVINETLRIGNVRFELHRRALKDVRKGYDIPSGVVLVISAIVHND 416
 DB 321 GLEDVSKMFTRAYIETSRILATVNGLRKTITDLENGLIRKGRITVYTTREINDA 380
 QY 417 SRYPQNLFNFRWQOONNGASSSGSFSSTWGNMYPFGGPRCAGSELAKLEMAVFI 476
 DB 381 NLVEDPLLFNFRWKKKSLESO-----NSCEVFGGTRLCPEKELGIVEISFL 429
 QY 477 HHLVLFNMEIAEDDOPFAFPVDFPNGLPIRVS 510
 DB 430 HYFTRVMEIGDELMAVFPFVAPAPGFIHLIS 463
 RESULT 12
 Q940V4 PRELIMINARY; PRT; 465 AA.
 AC Q940V4; PRELIMINARY; PRT; 465 AA.
 DT 01-DEC-2001 (TREMREL. 19, Created)
 DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
 DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
 DE A0330180/T20F20_6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carlini P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,

RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.,
 RT "Arabidopsis cDNA clones."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kim C.J., Chen H., Cheuk R., Meyers M.C., Banh J., Bowser L.,
 RA Carlini P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,
 RT "Arabidopsis ORF clones."
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AY063728; AAU36078.1; -
 DR EMBL: AY063728; AAU36078.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 465 AA; 53814 MW; 280A21D0712PCA11 CRC64;
 Query Match 26.6%; Score 712; DB 10; Length 465;
 Best Local Similarity 32.1%; Pred. No. 4.3e-43;
 Matches 163; Conservative 94; Mismatches 194; Indels 56; Gaps 8;
 QY 14 LPLSLSLFLILKRRN--KTRRFLPGKSGWPLGTTIGYLRPYATTGDMQ 67
 DB 3 IMMILGLIVITVLCRLALRNOMRYSKGLPGTGMWPLFGTTEFLKGP-----DF 57
 QY 68 MOHVSXKGIYRSNLEGEPTVSADAGLNRFILNNGRLFECSYPSISGILGKSMVLV 127
 DB 58 MNQRLRGSEFKSHLIGCPILISMDESVNRKILAKNSKGLVAGYPOSMDLIGTCMAAVH 117
 QY 128 LVGDHDMSSISLNTSHARLRTILKDYERHTLFYLDSSMOONSISADDEAKKFTF 187
 DB 118 VHGSHRLMKSLSLISPTMMRDHILPKYDHEFRSYLDQMLEVLDIDDKTHMAFLSL 176
 QY 188 MAKHIMSDPGEETEOLEKKEVYTFMGVAPLNPCTAYHKAQSRATILKFEIKME 247
 DB 177 SSLQIAETLKKPEVEYRTFEFLVGLISVPLIDIGTYRSGVQARRANNIDRLTELMQ 236
 QY 248 ERKLDKEEDOEKEEVKTEDEAEMSKSDHVRKORTDDLLGWVLKHSN--LSTEQILD 303
 DB 237 ERK-----ESGETP-----DMIGYLMKKEDNRYLLTDEKIRD 269
 QY 304 LLSLFGHETSSVAIALAIFFLQACRAVEERREHLEIARAKKEGSELAMNDYK 363
 DB 270 QVYTLISGYETVSTSMALIKYLHDKALQELRAELAFERERKQ--DEPLTIDIS 327
 QY 364 MDTQCVINETLRIGNVRFELHRRALKDVRKGYDIPSGVVLVISAIVHND 423
 DB 328 MKFTRAYIETSRILATVNGLRKTITDLENGLIRKGRITVYTTREINDA 387
 QY 424 LFNPMWQOONNGASSSGSFSSTWGNMYPFGGPRCAGSELAKLEMAVFIHLVLF 483
 DB 388 IFNPMWKKSSLSKS-----YFLLFGGVRLCPGKELGISEVSSFLHYFTVK 436
 QY 484 NMEIAEDDOPFAFPVDFPNGLPIRVS 510
 DB 437 RMEENGDELMAVFPFVAPAPGFIHLIS 463
 RESULT 13
 Q9LH81 PRELIMINARY; PRT; 465 AA.
 AC Q9LH81; PRELIMINARY; PRT; 465 AA.
 DT 01-DEC-2001 (TREMREL. 19, Created)
 DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
 DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
 DE A0330180/T20F20_6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carlini P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,

01-OCT-2000 (TREMblrel. 15, Created)
 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 Cytochrome P450.
 Arabidopsis thaliana (Mouse-ear cress).
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA:
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA:
 RA MEDLINE=20363099; PubMed=10907853;
 RX Nakamura Y.;
 RA "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones."
 RL DNA Res. 7:217-221(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL, AP002060, BAB02270.1, -.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PS00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 465 AA; 53862 MW; 630A21D0765EDDD CRC64;

Query Match 26.5%; Score 711; DB 10; Length 465;
 Best Local Similarity 32.1%; Pred. No. 5,1e-43;
 Matches 163; Conservative 93; Mismatches 195; Indels 56; Gaps 8;

0Y 14 LLLPSLLSLFLIL-----KRRNRKTRFNPGRSGMPFGITGYLKYPTATLGD 67
 DB 3 IMMMLGLLVITVICTALLRNMRKSKGLPPTMGMPFGITTEFLKQGP-----DE 57
 68 MOOHVSKYKTIYRSNLFCEPTIVSADAGLNRFILQNEGRLEFCSYPSISGILGKMSLV 127
 DB 58 MKNQGLRKGSPFKSHILGCPITVSMDAELNRYILNNEKSGLVAGYFQSLDLTGICNTAA 117
 0Y 128 LVGDNRDMRSTISLNFSLHARLITLKDVERHTLFLDSWQNSIFSAODEAKFTNL 187
 DB 118 VHGPHRMRGSLSLISPTMKDHLPKIDFMNRYLCGMDELTVDIQEKTKMAF-L 176
 0Y 168 MAKHIMSDPGEETEOLKKEVTFMKGVSAFLNPGTAYHKALOSATILKTERKME 247
 DB 177 SSLQIAETLKKPEVEERYKTEFFKLVGTLSPVLDIPGTNRSFGFOARNNDRLTTELMQ 236
 0Y 248 EKKLDIKEDDEEVEVKTEDEAKMSKSDHVRKQRTDDLLGVNLKSN---LSTEQILD 303
 DB 237 ERK-----ESGETFL-----DMLGVLKMKEDNRYILYLTKEIRD 269
 0Y 304 LILSLFPGHETSSVAIALAIFLQACRAVEELREHELELAKAKELGESELMNDYKK 363
 DB 270 QVVTLVGYEYVSTSMALKYLLHDHKALEELREHIALIREKRP--DEPLTLDIDKS 327
 0Y 364 MDFQCIVNETLRLGNVYRFLHRKALKDVRKGYDIPSGMKVLPVISAVHLDNSRYDDPN 423
 DB 368 MKETAVVIFERSRLATYNGVLRKTHDELLNGVLLIPKGMRYVYVTRINDTSLYEDPM 387
 0Y 424 LFNPRMOQNNAGSSSGSFSTGNVMPYGPRLCAGSELAKLEMAVYIHLVLKF 483
 DB 368 IENPRMKELESKS-----YFLEGGGVRLCPGKELGISVSSLHIFVYKY 436
 0Y 484 NMELAEDDQPFAPFVDPNGLPIRVS 510
 DB 437 RWEENGEDKLWFPFVSAPKGYHCKS 463

RESULT 14
 0949P1 PRELIMINARY; PRT; 467 AA.
 ID 0949P1;
 AC 0949P1;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Putative cytochrome P450 protein.
 GN ATG19230.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RP Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
 RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Kosemura E., Lam B.,
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full Length cDNA of gene ATG19230 (GI:7268718).";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RP Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AY050980; AAK3657.1; -
 DR EMBL: AY091446; AAM14385.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 467 AA; 53037 MW; 2F4230446536D955 CRC64;

Query Match 25.1%; Score 672.5; DB 10; Length 467;
 Best Local Similarity 32.3%; Pred. No. 3e-40;
 Matches 164; Conservative 96; Mismatches 183; Indels 65; Gaps 15;

0Y 13 LLLPSLLSLFLILKRRNRKTRFNPGRSGMPFGITGYLKYPTATLGDPMQHV 72
 DB 10 LEAGSLFLYFLRCLLSQRRFGSSKILPPTMGWYVGT---PQLYSQDP-NVFQSQ 65
 0Y 73 SKYGTIYSNLFCEPTIYSDAGLNRFILQNEGRLEFCSYPSISGILGKMSLVYGD 132
 DB 66 KRYGVFETHYVAGCPVVISPEAKFVLYTSKHLFFPFSKSRMRGKQAIFFHOGY 125
 0Y 133 HDKRSISLNFSLHARLITLKDVERHTLFLDSWQNSIFSAODEAKFTNLMAKI 192
 DB 126 HAKRLKLVLRAMPESIRN-WVPDIESTIADSLASW-BETMNTYQEMKTYTFNALLSI 183
 0Y 193 MSMDGEET---EOLKKEVTFMKGVSAFLNPGTAYHKALOSATILKTERKME 249
 DB 164 F---GKDEVLYRELKRCYIYLEKGYNSMPVNLGTLFKMKARKKELSOILAILISR 239
 0Y 230 KIDIKEDDEEVEVKTEDEAKMSKSDHVRKQRTDDLLG-VWLKSNSTEOIDLLISL 308
 DB 240 R-----ONGSSH-----NDLLGSFNGDKDEELTDEQIADNIIGV 272
 0Y 309 LEAGHETSSVAIALAIFLQACRAVEELREHELELAKAKELGESELMNDYKKMDFTQ 368


```

Db 273 IFARDTASVMSILKIYLAENPNVLEAITEQMAI-RKDEGES-LTWGDTKKMPLTS 330
QY 369 CYINETLRIGNVRLHRRALDVKYKGYDIPSGWKVLPVISAHLNDSRYDQPLEMPW 428
Db 331 RVIQETLRVASILSTFEAREAVDEYEGYILPKGKVLPLFRNIHSHADITSNPKFDP 390
QY 429 RMOQONNGASSSGSFSFTGNNYMPFGGPRLCAGSELAKLEMAVFTIHLVLFKNWL- 487
Db 391 RE-----VAPKPTNTEFMPFGNGTHSCPGNELAKLEMSIMITHLTKYSMSIV 437
QY 488 -AEDD---OPFAFPVDPNGLPIRVSR 511
Db 438 GASDGIQYGPALP-----QNGLPYILAR 461

RESULT 15
Q9FH76 PRELIMINARY; PRT; 463 AA.
AC Q9FH76;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cytochrome P450 (AT5g45340/K9E15_12).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty pl and YAC
RT clones.";
RL DNA Res. 7:31-63(2000).
RN 2;
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
RA Becker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -1-SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB020744; BAB10255.1; -
DR EMBL; AY065065; AAL57698.1; -
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KM Heme; Monooxygenase; oxidoreductase.
SQ SEQUENCE 463 AA; 5236 MW; CCD17293F53F812 CRC64;

Query Match 24.5%; Score 658; DB 10; Length 463;
Best Local Similarity 31.8%; Pred. No. 3.3e-38;
Matches 163; Conservative 95; Mismatches 184; Indels 70; Gaps 15;

QY 14 LLPSLLSLFLILK-----RRNRKTRFNLPPKSGMPFLGETIGYIKPYATATLGDPM 68
Db 6 LFTLLSAALFLCLRLFTAGYRRSSSTKLPLPPTMGYPYGER---FQLXSODP-NVFF 61
QY 69 QOHVSKYGTIRSNLFGPTIVSADAGLNFRIILONEGRLEFCSYPRSTIGILGKMSMLVL 128
Db 62 AAKQRRRGVSVEKTHVLCPCVMISPPAAKFLVLTAKSHLKPFTFPARKEMLKROAIFEH 121
QY 129 VGDHHRMRISLNLFLSHARLRLTLKDVHRHLFLVLDWQONSIFSADQEAKEFTFNLM 188

```

```

Db 122 QGDHSLKRLKLYLRAFPMDAIRN-MYPHISIAQESLNSMSDGLQWTFYQ-EKKTIFENVA 179
QY 189 AKHMSMDPGEERF---EOLKKEYVFMKGVSAPLNLPETAYHAKLSRAATLKEIRK 245
Db 180 LITL-----GKDEYTRIEDLKRCYILLEKGYNSMPNLPGTLFHKMKARKKELAQILANI 235
QY 246 MEERKLDIKREDDDEEEVKTEDEAEMSKSDPHVKQRKTDLDLGGWYK-HSNLSTQIIDL 304
Db 236 LSKRR-----QNPSSHT-----DLGSFWEKRAGLTDQIADN 268
QY 305 ILSLTAGHTSSVAIALAIFLQACPAAVEELREEHLEIARAKKELGSELNMDQYKM 364
Db 269 IIGVIFRAADPTASVLTWLKLYLADNPVLEAITEQMAIRKKE-GES-LTWGDTKKM 326
QY 365 DFTQCYINETLRIGNVRLHRRALDVKYKGYDIPSGWKVLPVISAHLNDSRYDQNL 424
Db 327 PLTYRVIQETLRVASILSTFEAREAVDEYEGYILPKGKVLPLFRNIHSHADITSNPKFDP 386
QY 425 FNPWRMOQONNGASSSGSFSFTGNNYMPFGGPRLCAGSELAKLEMAVFTIHLVLFKN 484
Db 387 FQPSRE-----VAPKPTNTEFMPFGNGTHSCPGNELAKLEISVLIHLLTKYR 433
QY 485 WEIAEDDQ-----PPAFPPVDPNGLPIRVSR 511
Db 434 WSIYVPSDGIQYGPALP-----QNGLPYILAR 461

```

Search completed: March 30, 2003, 12:06:27
Job time : 66 secs

•
•
•
•

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:06:10 ; Search time 10842.1 Seconds

(without alignments)
2955.362 Million cell updates/sec

Title: US-09-502-426a-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatcacaataatataatataat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY_NDC

Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank1:*

1: gb_ba:*

2: gb_hg:*

3: gb_in:*

4: gb_lm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hg_hum:*

31: em_hg_inv:*

32: em_hg_other:*

33: em_hg_mus:*

34: em_hg_pin:*

35: em_hg_rtd:*

36: em_hg_mam:*

37: em_hg_vrt:*

38: em_sy:*

39: em_hg_hum:*

40: em_hg_mus:*

41: em_hg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1101	100.0	4818	AF044216	AF044216 Arabidops
2	1090	99.0	84196	AT136	AT136 Arabidops
3	62.6	5.7	4629	AF273674	AF273674 Arabidops
4	61.4	5.6	19029	AE001372	AE001372 Plasmodiu
5	60.8	5.5	12930	AC117079	AC117079 Dictyoste
6	60.6	5.5	3364	TSP418778	TSP418778 Tomato sp
7	60.6	5.5	171317	AC020941	AC020941 Homo sapi
8	60	5.4	233269	AC094291	AC094291 Rattus no
9	59.4	5.4	224635	AL733314	AL733314 Homo sapi
10	59	5.4	253305	PFMAL3P7	PFMAL3P7 Plasmodiu
11	58.6	5.3	165669	AC114238	AC114238 Plasmodiu
12	58.6	5.3	192389	AC114238	AC114238 Plasmodiu
13	58	5.3	115489	AC007182	AC007182 Homo sapi
14	57.8	5.2	27291	AC115575	AC115575 Dictyoste
15	57.8	5.2	169546	AC004157	AC004157 Plasmodiu
16	57.6	5.2	35077	CEY38H8A	CEY38H8A Rattus no
17	57.6	5.2	144784	AC009224	AC009224 Homo sapi
18	57.6	5.2	188993	AC109636	AC109636 Dictyoste
19	57.2	5.2	18701	AC116985	AC116985 Dictyoste
20	57.2	5.2	18701	HSJ333B15	HSJ333B15 Human DNA
21	57.2	5.2	73666	CEY51H1A	CEY51H1A Caenorhab
22	57	5.2	43666	AC005504	AC005504 Plasmodiu
23	57	5.2	104992	AC005504	AC005504 Plasmodiu
24	56.8	5.2	46811	AC116550	AC116550 Dictyoste
25	56.8	5.1	160759	AC117082	AC117082 Dictyoste
26	56.6	5.1	6107	AX252043	AX252043 Sequence
27	56.6	5.1	6107	AX344427	AX344427 Sequence
28	56.6	5.1	6107	AX348832	AX348832 Sequence
29	56.6	5.1	56099	AC115598	AC115598 Dictyoste
30	56.6	5.1	67970	PFMAL1P3	PFMAL1P3 Plasmodiu
31	56.6	5.1	86515	AC127730	AC127730 Rattus no
32	56.6	5.1	159941	AC016220	AC016220 Homo sapi
33	56.6	5.1	178642	AC023122	AC023122 Homo sapi
34	56.4	5.1	65691	PFMAL3P1	PFMAL3P1 Plasmodiu
35	56.4	5.1	99003	AL390756	AL390756 Homo sapi
36	56.4	5.1	110000	PFMAL1P2_2	PFMAL1P2_2 Continuation (3 of
37	56.4	5.1	124041	CNS01DS1	CNS01DS1 Human chr
38	56.4	5.1	183353	AC087071	AC087071 Homo sapi
39	56.4	5.1	185273	AC073320	AC073320 Homo sapi
40	56.2	5.1	45296	AC115613	AC115613 Dictyoste
41	56.2	5.1	163443	AC006280	AC006280 Plasmodiu
42	56.2	5.1	196149	AC004709	AC004709 Plasmodiu
43	56	5.1	204439	AF000807	AF000807 Homo sapi
44	55.8	5.1	2166	AF200327	AF200327 Plasmodiu
45	55.8	5.1	14867	AE001398	AE001398 Plasmodiu

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AF044216	AF044216	Arabidopsis thaliana steroid 22-alpha-hydroxylase (DMF4) gene,	4818 bp	DNA	linear	PLN 25-JUN-2001	Arabidopsis thaliana.	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosoids II; Brassicales; Brassicaceae; Arabidopsids.	Choe,S., Dilkes,B.P., Fujioka,S., Takatsuto,S., Sakurai,A. and
AF044216	AF044216	Complete cds.					Arabidopsis thaliana.		
AF044216	AF044216	GI:2935341					Arabidopsis thaliana.		

TITLE Feldmann, K.A.
The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates multiple 2alpha-hydroxylation steps in brassinosteroid biosynthesis
JOURNAL Plant Cell 10 (2), 231-243 (1998)
MEDLINE 98158690
PUBMED 9490746
REFERENCE 2 (bases 1 to 4818)
AUTHORS Choe, S., Dikes, B.P., Azpiroz, R. and Feldmann, K.A.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-1998) Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
FEATURES
Source location/Qualifiers
1..4818
/organism="Arabidopsis thaliana"
/strain="Ws-2"
/db_xref="taxon:3702"
/chromosome="III"
/map="8 cm from nga6"
990..4398
/gene="DWF4"
/note="dwarf4"
/allele="dwarf4-1; dwarf4-2; dwarf4-3; dwarf4-4"
990..1055
/gene="DWF4"
1010(<1133..1353,1434..1758,1844..1996,2095..2409,2562..2634,2746..2824,2931..3040,3795..4398)
/gene="DWF4"
/product="steroid 22-alpha-hydroxylase"
1010(1133..1353,1434..1758,1844..1996,2095..2409,2562..2634,2746..2824,2931..3040,3795..4040)
/gene="DWF4"
/function="brassinosteroid biosynthetic enzyme"
/note="member of the cytochrome P450 superfamily; CYP90B1"
/codon_start=1
/product="steroid 22-alpha-hydroxylase"
/protein_id="AAC05093.1"
/db_xref="GI:2935342"
/translation="MFEERHTLPLLLPSLLSLFLILKRNKRNLPKPKS
GMPFLGRTIGLPTATATLADFMQOHVSKYKLYRSLFEPTIVSADGLNPTILQ
NEGRILPCSPRSIGLIGKWSMLVLDGMDHMRKSLISLNSHARLRTILKVERH
TLFLVDSMOQNSISADDEAKFTFNLMKIMSDGEETEDLKREYFMKGVVS
APLNLPGTAAVKALOSRATILKTEERKMEERLKDKEDEEVEKTEDEEMKSDH
VRKQRTDILLGWLKHNSLSTBOILLIILSLAGHETSVAIALAFTFGQAPKAY
EELREHLEIRAKRELGESELMNDYKKMDPTFCVINEIRLGNVAFELRKALDY
RYKGDIPSGKVLPIVSAVLNDSRYDQPNLEPMWQNNNGASSSGSGSFTWGN
NYMPFGGPRLCAGSELAKLEMAVFIHLVLKFWMELEDDQPEAFPEVDPNGLPIR
VSRIL"
BASE COUNT 1566 a 727 c 888 g 1637 t
ORIGIN
Query Match 100.0%; Score 1101; DB 8; Length 4818;
Best Local Similarity 100.0%; Pred. No. 4.3e-180;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATCACAATATATATATAGTCATATACATGATGAGAAAGTCCCAAAAAATTT 60
DB 32 AATCTCAAAATATATATATAGTCATATACATGATGAGAAAGTCCCAAAAAATTT 91
QY 61 TGTAAAGAACTCCAAATTTTTTTTTATGAGAACAGAAATTAACAGATAGAAAC 120
DB 92 TGTAAAGAACTCCAAATTTTTTTTTATGAGAACAGAAATTAACAGATAGAAAC 151
QY 121 TATTTTGTGTGGAATGAGTAAATATACATTAAGCAAAATTTTAAAAATTAATAT 180
DB 152 TATTTTGTGTGGAATGAGTAAATATACATTAAGCAAAATTTTAAAAATTAATAT 211
QY 181 GCCATAGCCCTCAAAATATATGTTATCTAGTAGGTATATATATATGATGATGCGATT 240
DB 212 GCCATAGCCCTCAAAATATATGTTATCTAGTAGGTATATATATATGATGATGCGATT 271
QY 241 CAGAAATGGAGCAACATGAAAGCAATTAATATTAATTAATTAATTAATTAATTAAT 300

DB 272 CAGAAATGGAGCAACATGAAAGCAATTAATATTAATTAATTAATTAATTAATTAAT 331
QY 301 TGAGTAATATGTTTTCCTGACTATTTAGAGGGCAAAAAAGCAATGCCAAAGTCTAGC 360
DB 332 TGAGTAATATGTTTTCCTGACTATTTAGAGGGCAAAAAAGCAATGCCAAAGTCTAGC 391
QY 361 GGTGACATGCTGAGTGGTATATATATATATATATATATATATATATATATATATAT 420
DB 392 GGTGACATGCTGAGTGGTATATATATATATATATATATATATATATATATATATAT 451
QY 421 AGGGGCTTCTGACATTTTCACTGTTTACCCCTCTGAGCCACCTTTTCCCAT 480
DB 452 AGGGGCTTCTGACATTTTCACTGTTTACCCCTCTGAGCCACCTTTTCCCAT 511
QY 481 ATCTAAGGATATATTTTGGAAATCCCAATTTAAACGATTTGAGCCGATCCGATCTCT 540
DB 512 ATCTAAGGATATATTTTGGAAATCCCAATTTAAACGATTTGAGCCGATCCGATCTCT 571
QY 541 GGGATTTCTGCTGAGCATTTATCAAAATATATATATATATATATATATATATATAT 600
DB 572 GGGATTTCTGCTGAGCATTTATCAAAATATATATATATATATATATATATATATAT 631
QY 601 ACTCAACATTTGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
DB 632 ACTCAACATTTGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 691
QY 661 TCTAATGATTTTCTTCTTCTACCAAGGTGATGAAAGTATATATATATATATATATAT 720
DB 692 TCTAATGATTTTCTTCTTCTACCAAGGTGATGAAAGTATATATATATATATATATAT 751
QY 721 CAATGATATATGAT 780
DB 752 CAATGATATATGAT 811
QY 781 TATTTTCAGCATGCGACCTTCTGCAACTTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 840
DB 812 TATTTTCAGCATGCGACCTTCTGCAACTTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 871
QY 841 ACTATTTAAAGAGCATTAAGAGGCAACAAAGTATATATATATATATATATATATATAT 900
DB 872 ACTATTTAAAGAGCATTAAGAGGCAACAAAGTATATATATATATATATATATATATAT 931
QY 901 GAAGTCATGTTTGTGTTAAGCTTAATTAAGAAATTTATATATATATATATATATATAT 960
DB 932 GAAGTCATGTTTGTGTTAAGCTTAATTAAGAAATTTATATATATATATATATATATAT 991
QY 961 TAACATTTATATTTCTGACTTCTTAAACCCCTCTTACAAACGAGAGTCCCTTTT 1020
DB 992 TAACATTTATATTTCTGACTTCTTAAACCCCTCTTACAAACGAGAGTCCCTTTT 1051
QY 1021 CAGTAGAAGTCCGATCCCAATCTTAAAGACAAAGCCATTAAGAGAGAGTGAAGTGA 1080
DB 1052 CAGTAGAAGTCCGATCCCAATCTTAAAGACAAAGCCATTAAGAGAGAGTGAAGTGA 1111
QY 1081 AG 1101
DB 1112 AG 1132
RESULT 2
LOCUS AT1335/c 84196 bp DNA linear PLN 26-JAN-2000
DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
ACCESSION AL132979
VERSION AL132979.2 GI:6782244
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 84196)
AUTHORS Blocker, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quettler, F. and

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Salanoubat, M.
Unpublished
2 (bases 1 to 84196)
Eu Arabidopsis sequencing project.
Direct Submission
Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mpi-biochem.mpg.de, mayet@mpi-biochem.mpg.de Project
Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue
Gaston Cremieux, BP191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
On Jan 27, 2000 this sequence version replaced gi:6434247.
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
location/Qualifiers

FEATURES
source
1. 84196
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
2983. 3338
/note="355 bp LTR"
3339. 7564
/note="LTR-transposon"
7565. 7922
/note="367bp LTR"
12339. 13197
/gene="T3A5.10"
join(12339..12665,12742..12803,12888..13055,13125..13197)
/note="similarity to cyclin-dependent kinase inhibitor
(ICK1), Arabidopsis thaliana, EMBL:AT094772.1"
/codon_start=1
/product="putative protein"
/protein_id="CAB62432.1"
/db_xref="GI:6561966"
/translation="MAVRRRRDVEVENCYTTTVKRRKMEERVLVSRITLSPV
QATNRGIVARNAGASERSVIVRRRSDPPVEQCIIEEDSVSCGTSSEKSR
KEFVDELENNDDRETETSWITDILNKSESNMDSVAVEDVESRRIRLSLHETV
12339..12665
/gene="T3A5.10"
/number=1
12666..12741
/gene="T3A5.10"
/number=1
12742..12803
/gene="T3A5.10"
/number=2
12804..12887
/gene="T3A5.10"
/number=2
12888..13055
/gene="T3A5.10"
/number=3
13056..13124
/gene="T3A5.10"
/number=3
13125..13197
/gene="T3A5.10"
/number=4
15858..16470
/gene="T3A5.20"
complement(join(15858..15901,16014..16470))
/note="similarity to predicted proteins, Arabidopsis
thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB62433.1"
/db_xref="GI:6561967"
/translation="MARLEHNSNMGARISFSNPFVIRSEKSNKSNINSSRSPS

MSADAFSVTDYSMLPADEIFLKGKILPKETSHVHRTLGEELLTEERGSVDGNTF
SLRPIILSSSSFSFGTWRELLGLKRTHVSKTKDKVNEEVLSDQHKIISGNATREC
QVADTR"
complement(15858..15901)
/number=1
complement(15902..16013)
/number=1
complement(16014..16470)
/number=2
18303..19931
/gene="T3A5.30"
complement(18303..19931)
/codon_start=1
/product="scarecrow-like 7 (SCL7)"
/protein_id="CAB62434.1"
/db_xref="GI:6561968"
/translation="MAYMCTDSCGNLMAINAOULIKOKQOQSOHQOQERDQPNPWN
PSFGTLPGSGFSDPROVNTNDPGFHPHLEHONNAVASEEFDSDWMSLLINGDAS
QTMDDPIYGHDPFVSFSPRLSPSTLNKNVDSDASQQLPPPPASTAISMSPSPSQ
HPPPPQPDQDLNQPITFAIHIDYARKPKPTKPDITIRIKESVSGDPIQRYGYFAE
ALSHKTESPSSSSSLDEFTLSYKTLNDADPYSKFAHLTANQALILEATQSNMHI
VDGIRQGIQWSMALQALATRSRSGKPTRIRIGIPAPSLGDSGPGSLATGRLDFA
AIDLDNFEFYPVLTPIOLNGSSFRVDPDEVIVNPMLELYLIDDEATATGRTALRIA
RSINPRTIVTIGEYEVSLNVEFANRKNLSREYSVAFESLEPNIDRDSKERIVARVL
FGRRIMDIVRSDDDNNKPGTRGILMEKEKQRTVLEKKGFEFPKPSNVAVSQAKLLM
NNYSYSLVSESRPFIISLANNVPLTLTVSSWR"
complement(18303..19931)
/number=1
26093..28999
/gene="T3A5.40"
complement(join(26093..26338,27091..27200,27308..27386,
27478..27570,27723..28037,28136..28288,28374..28698,
28779..28999))
/note="Contains Cytochrome P450 cysteine heme-iron ligand
signature AA455-464; Prokaryotic membrane lipoprotein lipid
attachment site AA320-330"
/codon_start=1
/product="steroid 22-alpha-hydroxylase (DWF4)"
/protein_id="CAB62435.1"
/db_xref="GI:6561969"
/translation="WFEHTHTLLPULLPSSLILFLILKKRRNRKTRNLPFGKS
GMPFGETIGYIKPYTATVLDLDFMOQHSKYGKIVRSNLFEPITVSADAGLNFTIO
NEGRLECSYSPRSIGILGKWSMLVLDGMDRDRSISLNTLSHARTITILKVEEH
TLFVDSWQNSIFSADDEAKKFTFNMAKHIMSDPGEERTDOLKKEYYTFKAGVS
APLNPGTAYHKALOSRAITILKFERKKEPKKIDIKEDDEEERKTEDEAKSKSDH
VRKORTDDILGWLKHSNLSSTEOILDIILSLRAGHETSSVALAIAIFLQACPKAV
BEIAREHEIARAKKELESSELNMDYKKMPTOCVINEITRLIGNVYFLHRAKAKV
RYKGDITPSGKVLPEVISAHLDNSRDPNLPFPMRWQOQNNGASSGSGSFTGNC
NYMPGGGPRCLACGSELAKLEMAVEFIHHLVLKFNWELAEDDKPPAFPFVDPNGLPIR
VSRIL"
complement(26093..26338)
/number=1
complement(26339..27090)
/number=1
complement(27091..27200)
/number=2
complement(27201..27307)
/number=2
complement(27308..27386)
/number=3
complement(27387..27477)
/number=3
complement(27478..27570)
complement(27571..27722)
/number=4
complement(27723..28037)
/number=5
complement(28038..28135)
/number=5
complement(28136..28288)
/number=6
complement(28289..28373)

	Query Match	Best Local Similarity	Matches 1101: Conservative	Score 99.0%	DB 8:	Length 84196;
exon	1	99.0%	99.9%	1090;	DB 8:	Length 84196;
intron	2	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	3	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	4	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	5	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	6	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	7	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	8	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	9	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	10	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	11	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	12	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	13	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	14	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	15	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	16	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	17	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	18	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	19	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	20	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	21	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	22	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	23	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	24	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	25	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	26	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	27	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	28	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	29	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	30	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	31	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	32	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	33	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	34	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	35	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	36	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	37	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	38	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	39	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	40	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	41	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	42	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	43	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	44	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	45	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	46	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	47	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	48	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	49	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	50	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	51	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	52	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	53	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	54	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	55	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	56	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	57	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	58	99.9%	99.9%	1090;	DB 8:	Length 84196;
exon	59	99.9%	99.9%	1090;	DB 8:	Length 84196;
intron	60	99.9%	99.9%	1090;	DB 8:	Length 84196;

Db	29741	GGTTGAGCTGCACAGTCGGTAATTAATCTAATTAATCTGTCTTTAGCCGACACGCTCGTGT	29682
Qy	421	AGGGGTCTCTTGACATTTTTCACGTCTTACCCCTACCTCGAGCCACACCTTTTCCAT	480
Db	29681	AGGGGTCTCTTGACATTTTTCACGTCTTACCCCTACCTCGAGCCACACCTTTTCCAT	29622
Qy	481	ATCCCAAGGGTAATTTTGGAAAATCCCAATTTTAAACCGAATTTGGACCGATACGGACTTCT	540
Db	29621	ATCCCAAGGGTAATTTTGGAAAATCCCAATTTTAAACCGAATTTGGACCGATACGGACTTCT	29582
Qy	541	GGGATTTGCTGGAGACATTTATCAAAAATTTATACACGAATGGGTTTATTAATTTAAA	600
Db	29561	GGGATTTGCTGGAGACATTTATCAAAAATTTATACACGAATGGGTTTATTAATTTAAA	29502
Qy	601	ACTCAACAATGATCAGATTAATAATTTCAATTAACACTTTTACGAGGATTCGTACATCTA	660
Db	29501	ACTCAACAATGATCAGATTAATAATTTCAATTAACACTTTTACGAGGATTCGTACATCTA	29442
Qy	661	TCTATGACTTTTCTTTTCTTACACGGGTGATGAAATTTATAGTACTTTATAGCCAGGA	720
Db	29441	TCTATGACTTTTCTTTTCTTACACGGGTGATGAAATTTATAGTACTTTATAGCCAGGA	29382
Qy	721	CAATGATTAATAGATATATTCATTAATCCATGATATTTATGATATTAATATAGCTGTTAAC	780
Db	29381	CAATGATTAATAGATATATTCATTAATCCATGATATTTATGATATTAATATAGCTGTTAAC	29322
Qy	781	TATTTGACGATCGACGCTTTCTGCAACTTTGTTTAAATTAAGAGTTTATAATAATAA	840
Db	29321	TATTTGACGATCGACGCTTTCTGCAACTTTGTTTAAATTAAGAGTTTATAATAATAA	29262
Qy	841	AGTATTAATAAGGAGATTAACGAGGACAAAGTAATTAACACGAGAAACAAAAGCCAT	900
Db	29261	AGTATTAATAAGGAGATTAACGAGGACAAAGTAATTAACACGAGAAACAAAAGCCAT	29202
Qy	901	GAAGCTCAATGGTGA-GTTTAAAGCTTAATTAAGAAATTTATTAATTTAATGACAGATG	959
Db	29201	GAAGCTCAATGGTGAATTTAAGCTTAATTAAGAAATTTATTAATTTAATGACAGATG	29142
Qy	959	ATTAACAATTAATTAATTTCTGACTCTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT	1019
Db	29141	ATTAACAATTAATTAATTTCTGACTCTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT	29082
Qy	1020	TCAGTAGAAGTCCTGATTTCCCAATCTTAAAGACAAAGCCATTAGAAAGGAAAGTAGTGA	1079
Db	29081	TCAGTAGAAGTCCTGATTTCCCAATCTTAAAGACAAAGCCATTAGAAAGGAAAGTAGTGA	29022
Qy	1080	GAGAGAGAGAGAACTACTCC 1101	
Db	29021	GAGAGAGAGAGAACTACTCC 29000	
RESULT 3			
AF273674			
LOCUS	AF273674	4629 bp DNA linear	INV 24-APR-2001
DEFINITION	Plasmidium falciparum DNA-dependent RNA polymerase (TRNP) gene, complete cds; nuclear gene for probable mitochondrial product.		
ACCESSION	AF273674		
VERSION	AF273674.1	GI:9857984	
KEYWORDS			
SOURCE	Plasmidium falciparum.		
ORGANISM	Plasmidium falciparum.		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.		
AUTHORS	1 (bases 1 to 4629)		
TITLE	L4.J., Muga,J.A., Cernakian,N., Cedergren,R. and Feagin,J.E.		
JOURNAL	Identification and characterization of a Plasmidium falciparum RNA polymerase gene with similarity to mitochondrial RNA polymerases		
MEDLINE	Mol. Biochem. Parasitol. 113 (2), 261-269 (2001)		
PUBMED	21192559		
REFERENCE	2 (bases 1 to 4629)		
AUTHORS	L4.J., Muga,J.A., Cernakian,N., Cedergren,R. and Feagin,J.E.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-MAY-2000) Seattle Biomedical Research Institute, 4		

BASE COUNT 4945 a 1191 c 943 g 4950 t
ORIGIN


```

/db.xref="GI:20429083"  

/translacion="MSKVKLTKRESIVALLIOGKDLEEEEDONLVANPFKTCFLENLDQ  

IKKSVISCLTFEPKNRSIMKVIKOSDFTGKITKTSTRIQIGTIDTFRRLDSMIRV  

RUVETENSENELNTIKRSIHPILIQYGLPDDAKSAIAMIGSLPLIASVSFE  

MISVVALIYODANTKRDIDPKKDYDEALQCTVYSKSAFEMNEDOVKKREYAII  

LSSNPNAKSIIAMGHYSETLNKRFEPGVKKQAKLELA"  

BASE COUNT      1002 a      563 c      489 g      1310 t  

ORIGIN  
  

Query Match          5.5%; Score 60.6; DB 14; Length 3364;  

Best Local Similarity 48.2%; Pred. No. 0.38;  

Matches 171; Conservative 0; Mismatches 184; Indels 0; Gaps 0;  
  

QY       1 AATCACAATTATTTAATTAGTCATPAACATGCATGCAAAAGTCCAAAAAATTT 60  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

Db       2100 AATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT 2041  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

QY       61 TGTTAACAAGAAGCTCCAAATTTTTTTTTTTTATGGAACAGAAATACAGATGAAAAC 120  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

Db       2040 AAAAAAAAAAGTAAAAAAAAATAGCAATTAATAATAATAATAATAATAATAA 1981  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

QY       121 TATTTTGTTGTGAATGAGTAGTAGTAATATACATTAAAGCAATTTTAAAAATATATAA 180  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

Db       1980 AATAAATAATATAAATAATAAATAATAAATAATAAATAATAAATAATAAATAATAA 1921  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

QY       181 GCCTATACGCCGCTCAAAGTAGTTATCTAGTAGGTGTATTAATATGATCGTGCGATT 240  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

Db       1920 AAATAAAAAAAAAAAAAAAAAAAAAAAAATGAAAAATAAAAATAAAAATAAAAAA 1861  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

-QY     241 CAGATTGGACAACAAATGAAAGCGATTAAATATTAATTAATTAATAATTT 300  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

Db       1860 TAAAAAATAAAAAATAAAAAATAAAAAATATAAAAAATAAAAAATAAAAATAA 1801  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

QY       301 TGAGTAATATGTGTTTCTGACTATTGAGGCGCAAAAAAAGACAAATGCCAAAGT 355  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  

Db       1800 ATAAAAATAAAAATAAAAACAAAAAACAANAACAAAAACAAAAACCCAATT 1746  

           ||| |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||  
  

RESULT 7  

LOCUS   AC020941/c              171317 bp    DNA             linear    PRI 01-JUN-2001  

DEFINITION Homo sapiens chromosome 5 clone CTD-231219, complete sequence.  

ACCESSION AC020941  

VERSION   AC020941.5 GI:14277273  

KEYWORDS HTG.  

SOURCE   Homo sapiens.  

ORGANISM Homo sapiens  

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  

REFERENCE  

AUTHORS 1 (bases 1 to 171317)  

TITLE   DOE Joint Genome Institute and Stanford Human Genome Center.  

JOURNAL Direct Submission  

REFERENCE Unpublished  

AUTHORS 2 (bases 1 to 171317)  

TITLE   DOE Joint Genome Institute.  

JOURNAL Direct Submission  

REFERENCE Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint  

AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  

TITLE   3 (bases 1 to 171317)  

JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.  

REFERENCE Direct Submission  

AUTHORS Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell  

TITLE   Drive, Walnut Creek, CA 94598, USA  

JOURNAL On Jun 1, 2001 this sequence version replaced gi:12830142.  

COMMENT Draft Sequence Produced by DOE Joint Genome Institute  

www.jgi.doe.gov  

Finishing Completed at Stanford Human Genome Center  

www-shgc.stanford.edu  

Quality: Phrap Quality >=40 99.7% of Sequence:  

Estimated Total Number of Errors Is 2.7.  

Location/Qualifiers  

1..171317  

/organism="Homo sapiens"
```


10040 10139: gap of unknown length
10140 11793: contig of 1654 bp in length
11794 11893: gap of unknown length
11894 13371: contig of 1478 bp in length
13372 13471: gap of unknown length
13472 14864: contig of 1393 bp in length
14865 14660: gap of unknown length
14661 16560: contig of 1496 bp in length
16561 17663: contig of 1103 bp in length
17664 19281: contig of 1518 bp in length
19282 19381: gap of unknown length
19382 21112: contig of 1731 bp in length
21113 21212: gap of unknown length
21213 22933: contig of 1621 bp in length
22934 24214: contig of 1281 bp in length
24215 24314: gap of unknown length
24315 25527: contig of 1213 bp in length
25528 25628: gap of unknown length
25629 27221: contig of 1593 bp in length
27221 27320: gap of unknown length
27321 28926: contig of 1606 bp in length
28927 29026: gap of unknown length
29027 30363: contig of 1337 bp in length
30364 30463: gap of unknown length
30464 31665: contig of 1202 bp in length
31666 33349: gap of unknown length
33350 33449: contig of 1584 bp in length
33450 36317: gap of unknown length
36318 36417: gap of unknown length
36418 37778: contig of 1361 bp in length
37779 39851: contig of 1973 bp in length
39852 39951: gap of unknown length
39952 42802: contig of 2851 bp in length
42803 42902: gap of unknown length
42903 45682: contig of 2780 bp in length
45683 45782: gap of unknown length
45783 49368: contig of 3486 bp in length
49369 51428: gap of unknown length
51429 51528: contig of 2060 bp in length
51529 54571: gap of unknown length
54572 54671: gap of unknown length
54672 57546: contig of 2875 bp in length
57547 57646: gap of unknown length
57647 60741: contig of 3094 bp in length
60741 60840: gap of unknown length
60841 63916: contig of 3076 bp in length
63917 64016: gap of unknown length
64017 67000: contig of 2984 bp in length
67001 67100: gap of unknown length
67101 70219: contig of 3119 bp in length
70220 70319: gap of unknown length
70320 72950: contig of 2631 bp in length
72951 73050: gap of unknown length
73051 76802: contig of 3752 bp in length
76803 76902: gap of unknown length
76903 79990: contig of 3088 bp in length
79991 80090: gap of unknown length
80091 82745: contig of 2655 bp in length
82746 82845: gap of unknown length
82846 87725: contig of 4880 bp in length
87726 87825: gap of unknown length
87826 91178: contig of 3353 bp in length
91179 91278: gap of unknown length
91279 94533: contig of 3255 bp in length
94534 94634: gap of unknown length
94635 99871: contig of 5238 bp in length
99872 99971: gap of unknown length

99972 103742: contig of 3771 bp in length
103743 103842: gap of unknown length
103843 108877: contig of 5035 bp in length
108878 108977: gap of unknown length
108979 114019: contig of 5042 bp in length
114020 114119: gap of unknown length
114120 119231: contig of 5112 bp in length
119232 119331: gap of unknown length
119332 125649: contig of 6318 bp in length
125650 132687: contig of 6938 bp in length
132688 132788: gap of unknown length
132789 138741: contig of 6954 bp in length
138742 139842: gap of unknown length
139843 146349: contig of 6508 bp in length
146350 155325: contig of 8876 bp in length
155326 155425: gap of unknown length

Query Match 5.4%; Score 60; DB 2; Length 233269;
Best Local Similarity 51.3%; Pred. No. 0.2;
Matches 138; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

6 ACAATTTATATATTGTCATATACATGCAATGGAAGTCCAAAAATTTGTTA 65
1391 AAAAAATTTTGTGTTTTTAAAGTTAAAAATTTTAAAAATTTTGT 1450
66 ACAGAACTCCAAATTTTTTTTATGCAACAGAAATACAGATGAAACTATTT 125
1451 TTTTAAAAAAATTTTGTGTTTTTAAAAATTTTAAAAATTTTATTT 1510
126 TGTGTGGAATGGAAGTATATATATATACAAATTTTAAAAATTTATTAAGCCTA 185
1511 TTTTAAAAAAATTTTAAAAATTTTGTGCAAAAAATTTTAAAAATTTT 1570
186 TACGGCTCAAGTGTATCTAGTACGTGTAATATATATGCTGCGATTCACAA 245
1571 TTTTAAAAAAATTTTAAAAATTTTGTGCAAAAAATTTTAAAAATTTT 1630
246 TTGGCAACAATGAAACGCAATTTAAA 274
1631 TTTTAAAAAAATTTTAAAAATTTTAAAA 1659

RESULT 9
AL732314 224635 bp DNA linear HTG 17-AUG-2002
LOCUS
DEFINITION Homo sapiens chromosome X clone RP13-465B17, *** SEQUENCING IN
PROGRAMS *** 8 unordered pieces.
ACCESSION AL732314
VERSION AL732314.10 GI:22415930
KEYWORDS HTG; HTGS-PHASE1; HTGS-ACTIVEFIN; HTGS-DRAFT; HTGS-FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 224635)
AUTHORS Whitehead,S.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 21, 2002 this sequence version replaced gi:22204483.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: DB465B17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads

5419	contlig of 1512
10930:	gap of unknown
10931	

3419	contig of 1512 bp in length
10930:	gap of unknown length
10931	

11031 13469: contig of 2439 bp in length
13470 13569: gap of unknown length
13570 14675: contig of 1106 bp in length
14676 14775: gap of unknown length
14776 16102: contig of 1327 bp in length
16103 16302: gap of unknown length
16303 18320: contig of 2118 bp in length
18321 18420: gap of unknown length
18421 20658: contig of 2238 bp in length
20659 22073: gap of unknown length
22074 22173: gap of unknown length
22174 24783: contig of 2610 bp in length
24784 24883: gap of unknown length
24884 27460: contig of 2577 bp in length
27461 27560: gap of unknown length
27561 29143: contig of 1583 bp in length
29144 29243: gap of unknown length
29244 31382: contig of 2138 bp in length
31382 31481: gap of unknown length
31481 33731: contig of 2250 bp in length
33732 33831: gap of unknown length
33832 36867: contig of 3036 bp in length
36868 36967: gap of unknown length
36968 39884: contig of 2917 bp in length
39885 39985: gap of unknown length
39985 42654: contig of 2670 bp in length
42655 42754: gap of unknown length
42755 44617: contig of 1863 bp in length
44618 44718: gap of unknown length
44718 46244: contig of 1527 bp in length
46245 46344: gap of unknown length
46345 49733: contig of 3389 bp in length
49734 49833: gap of unknown length
49834 52093: contig of 2260 bp in length
52094 52193: gap of unknown length
52194 54926: contig of 2733 bp in length
54927 55026: gap of unknown length
55027 58424: contig of 3398 bp in length
58425 58524: gap of unknown length
58525 61513: contig of 2989 bp in length
61514 61613: gap of unknown length
61614 63685: contig of 2072 bp in length
63686 63785: gap of unknown length
63786 66780: contig of 2995 bp in length
66781 66880: gap of unknown length
66881 70124: contig of 3244 bp in length
70125 70224: gap of unknown length
70226 74001: contig of 3777 bp in length
74002 74101: gap of unknown length
74102 77381: contig of 3280 bp in length
77382 77481: gap of unknown length
77482 80569: contig of 3088 bp in length
80570 80669: gap of unknown length
80670 83635: contig of 2966 bp in length
83636 83735: gap of unknown length
83736 87660: contig of 3925 bp in length
87661 87760: gap of unknown length
87761 90614: contig of 2854 bp in length
90615 90714: gap of unknown length
90715 95460: contig of 4746 bp in length
95461 95560: gap of unknown length
95561 100030: contig of 4470 bp in length
100031 100130: gap of unknown length
100131 106875: contig of 6745 bp in length
106876 106975: gap of unknown length
106976 111951: contig of 4976 bp in length
111952 112051: gap of unknown length
112052 116127: contig of 4076 bp in length
116128 116227: gap of unknown length
116228 120328: contig of 4101 bp in length
120329 120428: gap of unknown length
120429 124423: contig of 3995 bp in length

124424 124523: gap of unknown length
124524 132057: contig of 7534 bp in length
132058 132157: gap of unknown length
132158 139109 139108: contig of 6951 bp in length
139109 139208: gap of unknown length
139209 149165: contig of 9957 bp in length
149166 149266 149265: gap of unknown length
149266 165669: contig of 16404 bp in length.
Location/Qualifiers
1..165669
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-336N7"
BASE COUNT 48928 a 31280 c 30162 g 48184 t 7115 others
ORIGIN
Query Match 5.3%; Score 58.6; DB 2; Length 165669;
Best Local Similarity 53.8%; Pred. No. 0.38; Indels 2; Gaps 1;
Matches 142; Conservative 0; Mismatches 120;
QY 40 AGAAGTCCAAAAAATTTGTTAAGCAAACTCCAAATTTTATGTAAC 99
1457 AAAAAATTAATAATTAATAATAATAATAATAATAATAATAATAATA 1516
QY 100 AAGAATTAACAGTACAACTATTTGTTGGAATGCACTAGTAATACATTAAAC 159
1517 AAAAAAAAAAAAAAAAAATTTTGGGAAAAAAGGGAATAATTTTAA 1576
QY 160 AATTTTAAAAATTTATATAGCCGTATACGCGCTCAAGATATGTTATAGTGCTAA 219
1577 ATTATTTATTTAAGCATTAATAATAATAATAATAATAATAATAATA 1636
QY 220 TTAATTAATGATGTCGAT--CAGAAATGGACAAACATGAAATTAATAT 277
1637 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1696
QY 278 TAACCTTAATAATAATAATTT 301
Db 1697 AAAAAATTTTAATAATAATTT 1720
RESULT 12
AC007182 192389 bp DNA linear PRI 25-AUG-1999
LOCUS Homo sapiens chromosome 14 clone BAC 507E23 map 14q24.3, complete
DEFINITION
AC007182
ACCESSION AC007182
VERSION AC007182.3 GI:5708446
KEYWORDS HGC.
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 192389)
Rosen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S.,
Madan, A., Ratcliffe, A., Shafer, T. and Hood, L.
Sequencing of human chromosome 14
JOURNAL
REFERENCE 2 (bases 1 to 192389)
Young, J., Rosen, L., Madan, A., Qin, S., Abbasi, N., Dors, M.,
Dickhoff, R., James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S.,
Ratcliffe, A., Shafer, T. and Hood, L.
Direct Submission
JOURNAL
TITLE Submitted (30-MAR-1999) Multigabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
3 (bases 1 to 192389)
Rosen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S.,
Madan, A., Ratcliffe, A., Shafer, T. and Hood, L.
Direct Submission
JOURNAL
TITLE Submitted (25-AUG-1999) Multigabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA

1

GenCore version 5.1.4-p5-A578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:05:35 ; Search time 219.917 Seconds
(without alignments)
11274.505 Million cell updates/sec

Title: US-09-502-426a-1_COPY_2102_3202

Perfect score: 1101
Sequence: 1 aatcacaataattataat.....gagagagaagaactagctcc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : N_Geneseq.101002.*

```
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1101	100.0	6888	21	AAA59599	DNA encoding a cyt
2	56.6	5.1	6107	24	ABL70390	Chemically treated
3	56.6	5.1	6107	24	AA561342	Human gene regulat
4	56.6	5.1	6107	24	ABK31431	Signal transductio
5	54.2	4.9	875	21	AAA01920	Human colon cancer
6	54.2	4.9	6113	24	ABJ32803	Human immune syste
7	53.8	4.9	37973	24	ABJ34197	Human immune syste
8	53.8	4.9	83391	24	ABG67094	Human anglogenesis
9	53.4	4.9	17131	24	ABJ33053	Human immune syste

10	53.2	4.8	8170	24	ABK28257	DNA transcription
11	53	4.8	54108	24	ABK22782	Human high bone ma
12	53	4.8	57273	24	ABK22784	Human high bone ma
13	53	4.8	66933	22	ABA82625	Human HBM gene reg
14	53	4.8	72049	22	ABA82623	Human HBM gene reg
15	52.6	4.8	16228	24	ABL70459	Chemically treated
16	52.6	4.8	16228	24	AA561424	Human gene regulat
17	52.4	4.8	4501	21	ABK33968	Human DNA for stag
18	52.4	4.8	5940	21	AAA70105	Plasmodium falcipa
19	52.2	4.7	7746	24	ABK40047	Human chemically p
20	52.2	4.7	7746	24	ABL33856	Human immune syste
21	52.2	4.7	78925	21	AA899888	Human FN gene. Ho
22	52	4.7	5309	22	AA546527	Tumour suppressor
23	52	4.7	5309	22	ABK40039	Human chemically p
24	52	4.7	5309	24	ABL33736	Human immune syste
25	52	4.7	6636	24	ABN80023	Human chemically m
26	52	4.7	7459	24	ABK31382	Signal transductio
27	51.8	4.7	641	24	ABO56694	Human colon cancer
28	51.6	4.7	47108	24	ABK31511	Signal transductio
29	51.4	4.7	4041	21	AAA70170	Plasmodium falcipa
30	51.2	4.7	3738	21	AAA70178	Plasmodium falcipa
31	51.2	4.7	5690	22	AA545368	Chemically pretrea
32	51.2	4.7	5690	22	ABK28205	DNA transcription
33	51.2	4.7	23695	24	ABO66981	Human anglogenesis
34	51	4.6	10286	22	AA545309	Chemically pretrea
35	51	4.6	10286	24	ABK28148	DNA transcription
36	50.8	4.6	5880	22	AA546331	Tumour suppressor
37	50.8	4.6	5880	24	ABK28177	DNA transcription
38	50.8	4.6	6254	24	ABJ33621	Human immune syste
39	50.6	4.6	12405	22	AA545330	Chemically pretrea
40	50.6	4.6	12405	24	AA561143	Human gene regulat
41	50.6	4.6	12405	24	ABK28199	DNA transcription
42	50.4	4.6	5689	22	AA545384	Chemically pretrea
43	50.4	4.6	5689	22	AA546426	Tumour suppressor
44	50.4	4.6	5689	24	ABK28226	DNA transcription
45	50.4	4.6	8085	22	AA546479	Tumour suppressor

ALIGNMENTS

RESULT 1	AAA59599	standard; DNA; 6888 BP.
ID	AAA59599;	
AC	AAA59599;	
XX	14-NOV-2000 (First entry)	
DT		
XX	DNA encoding a cytochrome P450 enzyme designated DMF4.	
DE		
XX	DMF4, cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;	
KW	plant phenotype; cell elongation; ss.	
KM		
XX	Arabidopsis sp.	
OS		
XX		
EH	Key	Location/Qualifiers
FT	promoter	1..3203
FT		/*tag= a
FT	TATTA_signal	3060..3125
FT		/*tag= b
FT	CDS	3203..6110
FT		/*tag= c
FT		/product= "DMF4"
FT		/note= "contains introns"
FT	exon	3203..3423
FT		/*tag= d
FT	intron	3424..3503
FT		/*tag= e
FT	exon	3504..3828
FT		/*tag= f
FT	intron	3829..3913
FT		/*tag= g

D	b	2222	TATTTCTGTGGAAATGGAAGTAGTAAGTAAATACATTBAGCAATTTTAAAAAATAATATATAA	2281
O	y	181	GCCATATCGGGCCTCAAAGATGTATCTATCTAGTGAGTGATAATTAATCATGTGGCGATT	240
D	b	2282	GCCATATCGGGCCTCAAAGTATGTATCTAGTAGTGATAATTAATCATGTGGCGATT	2341
O	y	241	CAGAATTTGGGCACAACATGAAAAGGGAAATTTAAATTTACTTTTAAATTAATTAATTAAT	300
D	b	2342	CAGAATTTGGGCACAACATGAAAAGGGAAATTTAAATTTACTTTTAAATTAATTAATTAAT	2401
O	y	301	TGAGTAATATGTATTTCTTCGACTATTTGAGGGGCAAAAAAAAAAGCAATGCCAAAGCTACG	360
D	b	2402	TGAGTAATATGTATTTCTTCGACTATTTGAGGGGCAAAAAAAAAAGCAATGCCAAAGCTACG	2461
O	y	361	GCTTTGACTGTCCGACTGGTAAATATCTAATTAATCTGTCTTTTGACCAGCGCTCGTGT	420
D	b	2462	GGTTTGACTGTCCGACTGGTAAATATCTAATTAATCTGTCTTTTGACCAGCGCTCGTGT	2521
O	y	421	AGGGGCTCTTCTGACATTTTCACTGTCTTACCCCTACTCGGAGCCACCCCTTTCCCAT	480
D	b	2522	AGGGGCTCTTCTGACATTTTCACTGTCTTACCCCTACTCGGAGCCACCCCTTTCCCAT	2581
O	y	481	AATCTTAAGGGTAATTTTGGAAATCCCAATTTTAAACCGATTAGAACCCGATCCGGACTTCT	540
D	b	2582	AATCTTAAGGGTAATTTTGGAAATCCCAATTTTAAACCGATTAGAACCCGATCCGGACTTCT	2641
O	y	541	GGGATTTCTGCTGAGCATTTATTCAAAATTAATTAACAGAAATGGTATTAATTTAAAA	600
D	b	2642	GGGATTTCTGCTGAGCATTTATTCAAAATTAATTAACAGAAATGGTATTAATTTAAAA	2701
O	y	601	ACTGACAACCTGATCAGATTTAAATTTTCAATTAACCTTTTTCAGATGGATTCGATCATCTA	660
D	b	2702	ACTGACAACCTGATCAGATTTAAATTTTCAATTAACCTTTTTCAGATGGATTCGATCATCTA	2761
O	y	661	TCTAATGACTTTTTTTTTTTTCTACACAGGTGATGAAGTTATAGTACTTTAGCCAGAGA	720
D	b	2762	TCTAATGACTTTTTTTTTTTTCTACACAGGTGATGAAGTTATAGTACTTTAGCCAGAGA	2821
O	y	721	CAATTTGATTAAGATATATCCATTATATCCATGATATTTATGATATTAATAGCTGTTAAAC	780
D	b	2822	CAATTTGATTAAGATATATCCATTATATCCATGATATTTATGATATTAATAGCTGTTAAAC	2881
O	y	781	TATTTTCAGATCGCAGCTTTCTGCAACTTTTGTTTTAATTTAAGTTTAAATTAATTA	840
D	b	2882	TATTTTCAGATCGCAGCTTTCTGCAACTTTTGTTTTAATTTAAGTTTAAATTAATTA	2941
O	y	841	AGTATTTAAAGACGATACGAGGGAACAAAGTAATGAACACGAGAAACAAAGCCAT	900
D	b	2942	AGTATTTAAAGACGATACGAGGGAACAAAGTAATGAACACGAGAAACAAAGCCAT	3001
O	y	901	GAGGCTCATTTGTTAGTTTAACTTAATPACAGATTTTATTAATTTTATGAGCATGA	960
D	b	3002	GAGGCTCATTTGTTAGTTTAACTTAATPACAGATTTTATTAATTTTATGAGCATGA	3061
O	y	961	TAAATTAATATTTCTGACTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT	1020
D	b	3062	TAAATTAATATTTCTGACTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT	3121
O	y	1021	CAGTAGAAGTCCGATTTCCCATTTTAAAGCAAAAGCATTTAGAAAGAAAGTGAAG	1080
D	b	3122	CAGTAGAAGTCCGATTTCCCATTTTAAAGCAAAAGCATTTAGAAAGAAAGTGAAG	3181
O	y	1081	AGAGAGAGAGAACTAGCTAC 1101	
D	b	3182	AGAGAGAGAGAACTAGCTAC 3202	

RESULT 2
 ABL/0390/C
 ID ABL/0390 standard; DNA: 6107 BP.
 AC ABL/0390;
 XX

	Query Match	5.1%;	Score 56.6;	DB 24;	Length 6107;	
	Best Local Similarity	47.8%;	Pred. No. 0.044;			
	Matches 164;	Conservative 0;	Mismatches 179;	Indels 0;	Gaps 0	
OY	26 CAAATACAAATGATAGAAAAGTCCCAAAAAAATTTGGTTACAGAAACTCTCCAAATTCTT	85				
Db	4771 CAAAAAATAAATAAATAATTAATAATAAAAAATCATAAAATATCATTAATTAATTCCTT	4712				
OY	86 TTTTTTATGGAACAAGAATAAACAGATAGAAAACATATTTTGTTGTGGAAATGGAATAGT	145				
Db	4711 TTAAATTAATAAACTAAAAAACCACATAAAAATTTATATATACAAAAATCAAAAAATTTT	4652				
OY	146 AATATACATTAAGCAAATTTTAAAAATATATTAAGCCCTATFAGCGCGCTCAAGATATGTTA	205				
Db	4651 AAACCTACTATATATAAACTCATATATACATAAAAACGTAAAAAATTAACATAAATAAAA	4592				
OY	206 TCTAGTAGGTGAATTAATATATATGATGGTGCAGATTCAGATTGGGCAACAATGAAAACG	265				
Db	4591 TAAATTAATAAAAAATAATTAATTCATATATCATTAACAAAAATACATTAATCTATTAAT	4532				
OY	266 GAATTAATAATATTAACCTTTAAAAATTAATAAAAATTTGAGTAATGTGTTCTGACTAAT	325				
Db	4531 TAAAACTCATATATTTAAAAAATAACGAAAACTACTATTTATATACACTTTCTTTGCC	4472				
OY	326 GAGGGCAAAAAAAGACAAATGCCAAAAAGCTACGGGTTTGAC	368				
Db	4471 TTTAATCAAAAATTAATCATTTCCAAAAAATTAATCATTTTATAC	4429				
RESULT 4						
ABK31431/C						
ID	ABK31431 standard; DNA; 6107 BP.					
XX AC	ABK31431;					
XX DT	23-APR-2002 (first entry)					
XX DE	Signal transduction associated gene modified complementary DNA #137.					
XX KW	Human; signal transduction associated gene; cytosine methylation state;					
XX KM	Cpg Island; signal transduction associated disease; solid tumour; cancer;					
XX OS	antitumour; cytostatic; mutant; ds.					
XX OS	Homo sapiens.					
XX OS	Synthetic.					
PN PN	WO200200926-A2.					
PD PD	03-JAN-2002.					
PF PF	29-JUN-2001; 2001WO-EP07472.					
XX XX	30-JUN-2000; 2000DE-1032529.					
PR PR	01-SEP-2000; 2000DE-1043826.					
PA PA	(EPIC-) EPIDENOMICS AG.					
PI PI	Olek A, Pienbrock C, Berlin K;					
PT PT	WPI; 2002-147896/19.					
PS PS	Oligonucleotide for diagnosis and therapy of diseases associated with					
XX XX	signal transduction e.g. cancer, comprises chemically modified genomic					
XX XX	sequences of genes associated with signal transduction -					
XX XX	Claim 1; SEQ ID NO 274; 24pp; English.					

[illegible]

DT 26-MAR-2002 (first entry)

AC ABL34197;
XY

[illegible]

Query	Best Local Similarity	4.9%	Score 53.8	DB 24	Length 83391
Matches 165	Conservative 0	Mismatches 167	Indels 1	Gaps 1	
9	AATATTAATTAATTTAGTCATATACATGATGCAAGAGTTCACAAAAAATTTGTAAACA	68			
74888	ATTATTTGGGATTTATTTTAAAGAAATATATTAAGAAAGGTAATTAATTAATTAATTA	74947			
69	GAACTTCCAAATTTTATTTTATTTTATTTGGAACAAGAAATACAGATAGAAACTATTTGT	128			
74948	AATATTTTATTTGATTTATTTATTTATTTAGTAAAAATTTAGAAATTTTAAATATATAT	75007			
129	TGTGGAATGGAAGTGTATATTTACATTTACGAAATTTTAAAAATTTATATATAGCTATAC	188			
75008	AGTAG-ACAAAATGATTCAGTAATATATATATATATATATTTATATATATAGATATGGAAG	75066			

KW Human; mouse; Zmax1; HBM; high bone mass gene; lipid regulation; stroke;
KW lipid-associated condition; arteriosclerosis; cardiovascular disease; ss;
KW osteoporosis; atherosclerosis; diabetic atherosclerosis; plaque build-up;
KW neurovascular condition; wound healing; gene therapy; PCR primer; probe;
KW bone development disorder; antiarteriosclerotic; cardiovascular;
KW osteopathic; cerebroprotective.
XX
OS Homo sapiens.
PN WO200192891-A2.
XX
PD 06-DEC-2001.
XX
PF 25-MAY-2001; 2001WO-US16946.
XX
PR 26-MAY-2000; 2000US-0578900.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX (UYCR-) UNIV CREIGHTON SCHOOL MEDICINE.
XX
PI Carulli JP, Little RD, Recker RR, Johnson ML;
XX
DR WPI; 2002-097784/13.
XX
PT Identifying molecules involved in lipid regulation, useful for
PT diagnosing, treating or preventing e.g., arteriosclerosis, comprises
PT identifying a molecule that binds to high bone mass gene or its
XX corresponding wild type gene -
XX
PS Example 2; Page 350-392; 409pp; English.
XX
CC The invention relates to a method for identifying a molecule involved in
CC lipid regulation comprising identifying a molecule that binds to or
CC inhibits binding of a molecule to high bone mass (HBM) or its wild type
CC gene, Zmax1. Compounds identified by the method are useful for treating,
CC diagnosing, preventing or screening for normal and abnormal
CC lipid-associated conditions, including arteriosclerosis, cardiovascular
CC disease, stroke, and osteoporosis. The compounds may also be used in the
CC treatment or prevention of diabetic atherosclerosis, neurovascular
CC conditions caused by plaque build-up, poor circulation due to plaque
CC build-up and associated poor wound healing. The methods may be used in
CC gene therapy, pharmaceutical development, and diagnostic assays for bone
CC development disorders. Molecules identified by comparison of Zmax1 and
CC HBM systems can be used as surrogate markers in pharmaceutical
CC development, in diagnosis of human or animal bone disease, and in the
CC treatment of bone diseases. Sequences ABK22776-ABK23411 represent cDNA
CC molecules encoding human Zmax1 and HBM, and PCR primers, probes, linkers
CC and adapters of the invention.
XX
SQ Sequence 57273 BP; 12179 A; 15326 C; 15607 G; 14161 T; 0 other;
Query Match 4.8%; Score 53; DB 24; Length 57273;
Best Local Similarity 50.2%; Pred. No. 0.31;
Matches 157; Conservative 0; Mismatches 155; Indels 1; Gaps 1;

DB 16174 TAATATACATTTTAAATACACATTTATATTTATATATAAATATATAAATCTCC 16233
QY 305 TAAATGTGTTTC 317
DB 16234 AAGTCTCTTTTC 16246
RESULT 13
ABA82625
ID ABA82625 standard; DNA; 66933 BP.
XX
AC ABA82625;
XX
DT 25-JAN-2002 (first entry)
XX
DE Human HBM gene region b200e21-h_contig4.
XX
KW Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
KW sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
KW antisense therapy; vaccine; bone disorder; Paget's disease;
KW sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX
OS Homo sapiens.
XX
PN WO200177327-A1.
XX
PD 18-OCT-2001.
XX
PF 21-JUN-2000; 2000WO-US16951.
XX
PR 05-APR-2000; 2000US-0543771.
XX 05-APR-2000; 2000US-0544398.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Carulli JP, Little RD, Recker RR, Johnson ML;
XX
DR WPI; 2001-657171/75.
XX
PT New high bone mass (HBM) and Zmax1 genes and proteins useful for
PT modulating bone mass for the treatment of e.g. osteoporosis -
XX
PS Claim 51; Page 308-350; 443pp; English.
XX
CC The present invention describes the human Zmax1 gene and the high bone
CC mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and
CC HBM genes have osteopathic activities. The genes can be used in gene
CC therapy, antisense therapy and in the production of vaccines. They
CC can be used in the diagnosis and treatment of bone disorders including
CC osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
CC dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 66933 BP; 14237 A; 17817 C; 18323 G; 16556 T; 0 other;
Query Match 4.8%; Score 53; DB 22; Length 66933;
Best Local Similarity 50.2%; Pred. No. 0.32;
Matches 157; Conservative 0; Mismatches 155; Indels 1; Gaps 1;

DB 27154 ACATACCTTATAGTATATATTAATAATATATGTAATATTTTATTTATATGATATATA 27213
QY 245 ATTGGACACATGAAACGGAATTAAATATTAACCTTTAAATTAATTAATAATTGGAG 304
DB 27214 TAATATACATTTTAAATACATTTTATTTTATATATAATAATATATAATCTCC 27273
QY 305 TAAATGTGTTTC 317
DB 27274 AAGTTGCTTTTC 27286

RESULT 14
ABA82623
ID ABA82623 standard; DNA; 72049 BP.
XX ABA82623;
AC
XX
XX
DT 25-JAN-2002 (first entry)
DE Human HBM gene region b527812-h_contig309G.
XX
XX Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
KW sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
KW antisense therapy; vaccine; bone disorder; Paget's disease;
KW sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX
XX Homo sapiens.
XX
XX WO200177327-A1.
XX
XX PD 18-OCT-2001.
XX
XX PF 21-JUN-2000; 2000WO-US16951.
XX
XX PR 05-APR-2000; 2000US-0543771.
XX PR 05-APR-2000; 2000US-0544398.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX PA Carulli JP, Little RD, Recker RR, Johnson ML;
XX
XX PI WPI; 2001-657171/75.
XX
XX DR
XX
XX PT New high bone mass (HBM) and Zmax1 genes and proteins useful for
XX PT modulating bone mass for the treatment of e.g. osteoporosis -
XX
XX PS
XX PS Claim 51; Page 258-302; 443pp; English.
XX
XX CC The present invention describes the human Zmax1 gene and the high bone
XX CC mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and
XX CC HBM genes have osteopathic activities. The genes can be used in gene
XX CC therapy, antisense therapy and in the production of vaccines. They
XX CC can be used in the diagnosis and treatment of bone disorders including
XX CC osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
XX CC dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent
XX CC sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 72049 BP; 15164 A; 19306 C; 20142 G; 17434 T; 3 other;

Query Match 4.8%; Score 53; DB 22; Length 72049;
Best Local Similarity 50.2%; Pred. NO. 0.32; Indels 1; Gaps 1;
Matches 157; Conservative 0; Mismatches 155;

QY 6 ACAATATTAATATTAATGTCATTAACAATGCAATGAAAGTCCAAAAATTTGTTA 65
DB 30562 ATAAATATTAATATTAATTAATAATATTAATTAATTAATTAATTAATTAATA 30621
QY 66 ACAGAACTTCGAATTTTATTTTATGCA-ACAAGAAATACAGATACAAAATCTATT 124
DB 30622 AATAAATTTTAAATTTTATTTTATTAATTAATTAATTAATTAATTAATAATA 30681
QY 125 TTGTTGCAATGGAAGTAAATATACATTAAGCAAAATTTTAAAAATTAATATAGCT 184

DB 30682 TTATATATATATTAATTAATAATATATATATTAATAATATATATATATAT 30741
QY 185 ATACGGCGCTCAAGATGTTATCTAGTACTAGTGAATTAATTAATGATGCGATTGAGA 244
DB 30742 ACATACCTTATAGTATATATTAATAATATATGATATATTTTATATGATATATA 30801
QY 245 ATTGGACACATGAAACGGAATTAAATATTAACCTTTAAATTAATTAATAATTGGAG 304
DB 30802 TAATATACATTTTAAATACATTTTATTTTATATATAATAATATATAATCTCC 30861
QY 305 TAAATGTGTTTC 317
DB 30862 AAGTTGCTTTTC 30874

RESULT 15
ABL70459/C
ID ABL70459 standard; DNA; 16228 BP.
XX ABL70459;
AC
XX
XX
DT 01-JUL-2002 (first entry)
DE Chemically treated cell signalling DNA sequence#175.
XX
XX Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.
XX
XX OS
XX OS Unidentified.
XX
XX OS WO200202807-A2.
XX
XX PD 10-JAN-2002.
XX
XX PF 29-JUN-2001; 2001WO-EP07471.
XX
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX
XX PA (EPIC-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K;
XX
XX DR WPI; 2002-154758/20.
XX
XX PT Nucleic acid, useful for diagnosis and therapy of diseases associated
XX PT with cell signalling e.g. cancer, comprises chemically modified genomic
XX PT sequences of genes associated with cell signalling -
XX
XX PS
XX PS Claim 1; SEQ ID NO 349; 24pp+sequence listing; English.

CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligonucleotides for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.

XX
XX SQ Sequence 16228 BP; 4652 A; 198 C; 3601 G; 7777 T; 0 other;

Query Match 4.8%; Score 52.6; DB 24; Length 16228;
Best Local Similarity 48.2%; Pred. NO. 0.31; Indels 0; Gaps 0;
Matches 148; Conservative 0; Mismatches 159;


```

QY 47 TCCAAAAAATTTGTTAACAGAACTTCCAAATTTTTTTTTTTATGGAACAGAAAT 106
Db 7878 TCTACTATTAAACCTTATTTTAAAAAAACAACTTTTCTTTTAAACGAAATA 7819
QY 107 AACAGATAGAAACTATTTTGTGGAATGGAAGTAAATATACATTAAAGCAATTTT 166
Db 7818 TACCGTAACTAATTTATTTCCCAATTTCTACTTAATACCTCAATACCTAATAAATAACA 7759
QY 167 AAAAAATTAATATAGCCCTATACGGCTCAAGTATGTTATCTAGTGTAAATTAATA 226
Db 7758 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 7699
QY 227 TGCATGTCGATCAGATTCAGATTCGACAACTGAAGAAATTAATTAATTAATTA 286
Db 7698 TAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7639
QY 287 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 346
Db 7638 AACTATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7579
QY 347 GCCAAAA 353
Db 7578 TAAATAA 7572

```

Search completed: March 29, 2003, 18:32:56
 Job time : 644.917 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:09:40 ; Search time 38.5031 Seconds

(Without alignments)
8769.450 Million cell updates/sec

Title: US-09-502-426A-1_COPY_2102_3202

Sequence: 1 aatcacaattataat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents.NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.2	4.5	3350	US-08-617-860B-3	Sequence 3, Appl
2	49.2	4.5	6265	US-09-129-112-3	Sequence 3, Appl
3	46.6	4.2	19124	US-08-487-826B-13	Sequence 13, Appl
4	46.4	4.2	636	US-08-998-416-1137	Sequence 1137, Ap
5	46	4.2	731	US-08-451-405A-2	Sequence 2, Appl
6	45.8	4.2	665	US-08-883-795A-36	Sequence 36, Appl
7	45.8	4.2	51952	US-08-947-823-1	Sequence 1, Appl
8	45	4.1	20674	US-07-641-638-651	Sequence 651, App
9	44.8	4.1	3138	US-07-867-106-4	Sequence 4, Appl
10	44.6	4.1	19124	US-08-487-826B-13	Sequence 13, Appl
11	44	4.0	7218	US-08-232-463-14	Sequence 14, Appl
12	43.8	4.0	6124	US-08-213-419B-3	Sequence 3, Appl
13	43.6	4.0	5852	US-07-867-106-2	Sequence 2, Appl
14	43.2	3.9	660	US-07-991-867B-32	Sequence 32, Appl
15	43.2	3.9	660	US-08-107-755A-32	Sequence 32, Appl
16	43.2	3.9	660	US-08-544-332-32	Sequence 32, Appl
17	43.2	3.9	660	US-07-991-867B-32	Sequence 32, Appl
18	43.2	3.9	1511	US-09-370-861A-32	Sequence 8, Appl
19	43.2	3.9	1511	US-08-107-755A-8	Sequence 8, Appl
20	43.2	3.9	1511	US-08-544-332-8	Sequence 8, Appl
21	43.2	3.9	1511	US-09-370-861A-8	Sequence 8, Appl
22	43.2	3.9	4810	US-08-852-629-11	Sequence 11, Appl
23	43.2	3.9	4838	US-08-852-629-15	Sequence 15, Appl
24	42.8	3.9	1850	US-08-617-860B-32	Sequence 32, Appl
25	42.8	3.9	4098	US-08-605-106-4	Sequence 4, Appl
26	42.4	3.9	470	US-09-020-956-102	Sequence 102, App
27	42.4	3.9	470	US-09-030-607-102	Sequence 102, App

28	42.4	3.9	470	US-09-605-785-102	Sequence 102, App
29	42.4	3.9	470	US-09-439-313-102	Sequence 102, App
30	42.4	3.9	470	US-09-352-616A-102	Sequence 102, App
31	42.4	3.9	470	US-09-232-149A-102	Sequence 102, App
32	42.4	3.9	615	US-08-998-416-186	Sequence 186, App
33	42.4	3.9	5181	US-08-257-073-10	Sequence 10, Appl
34	42.2	3.8	1406	US-08-913-842-6	Sequence 10, Appl
35	41.8	3.8	658	US-08-998-416-595	Sequence 595, App
36	41.4	3.8	2861	US-08-299-953-1	Sequence 1, Appl
37	41.4	3.8	2861	US-08-459-415-1	Sequence 1, Appl
38	41.4	3.8	2861	US-09-066-687-1	Sequence 1, Appl
39	41.4	3.8	2861	PCT-US95-11231-1	Sequence 1, Appl
40	41.4	3.8	3881	US-08-239-953-2	Sequence 1, Appl
41	41.4	3.8	3881	US-08-459-415-2	Sequence 2, Appl
42	41.4	3.8	3881	US-09-066-687-2	Sequence 2, Appl
43	41.4	3.8	3881	PCT-US95-11231-2	Sequence 2, Appl
44	41	3.7	1441	US-08-821-994-63	Sequence 63, Appl
45	41	3.7	2110	US-09-419-459-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-08-617-860B-3/C

Sequence 3, Application US/08617860B

Patent No. 6133506

GENERAL INFORMATION:

APPLICANT: Tytfer, R., Bautor, J., Bothmann, H., Filisak, E.,

APPLICANT: Hytke-Grandpierre, C., Riello, B., Martini, N.,

APPLICANT: Moller, A., Schulte, W., Voeltz, M., Walek, J.,

APPLICANT: Schell, J.

TITLE OF INVENTION: Promoters

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESSES:

ADDRESS: Steuberg, Raskin & Davidson, P.C.

STREET: 1140 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/617,860B

FILING DATE: 01-MAR-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP94/02950

FILING DATE: 05-SEP-1994

APPLICATION NUMBER: DE P4329951.2

FILING DATE: 04-SEP-1993

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3350 base pairs

TYPE: Nucleic acid

STRANDEDNESS: Double stranded

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Brassica napus

IMMEDIATE SOURCE:

LIBRARY: genomic Lambda FIX II

CLONE: Baccaseg10

FEATURE:

NAME/KEY: Startcodon

LOCATION: 2611..2613

FEATURE:

NAME/KEY: CDS

```

? LOCATION: join(2611..2908, 3001..3341)
US-08-617-8608-3

Query Match      4.5%; Score 49.2; DB 3; Length 3350;
Best Local Similarity 48.7%; Pred. No. 0.02;
Matches 190; Conservative 0; Mismatches 198; Indels 2; Gaps 2

OY 6 ACAAAATTATTAATTTTGTGCAATTAACAATGCATAGAAAGTCCAAAAAATTTTGTGA 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2312 ATAAATGAACTTAATTTTATATATATATATATAT-CAGCTATGTAAATTAATATAGAA 2254
OY 66 ACAGAAACTTCCAAATTTTTTTTTTTTTTATGCAACAAGAAATTAACAGATAGAAACTATTT 125
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2253 ACTAAATTTATTAATTAATTAATGTTTTTAAACAATAATATAGTTGGGTA-TGTAT 2195
OY 126 TGTGTGGAAATGGAATGTATATACATTAAGCAAAATTTTAAAAAATTAATTAAGCCCTA 185
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2194 ATATTTTATGTTTAAAAACATTTACAAAAATATTTAATTTTAACTTTTAAATTTTATTTTC 2135
OY 186 TACGGCGTCAAAAGTATGTAATCTAGTAGTGTAATTAATTAATGCAATGGCGCATTCAGAA 245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2134 ATTTGACATATCAAAATATCAAAAAAATATGTAAGACTATTTTAAATTTTCTTCAACAAGT 2075
OY 246 TTGGAGCAACAATGAAACGCAATTAAATATTAATTAATTAATTAATTAATTAATTTGACT 305
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2074 AAGATTTATGATATATTTAAAAAATCAAGCATAAATTTATTAATTAATTAATTTAAT 2015
OY 306 AAATGTGTTTCTGACTATTGAGGGGCAAAAAAAGACAATGCCAAAAGCTACGGGTTT 365
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2014 GTGAATTAATTAATTAATTAACCTTTAAATATCAAAAAATTAACCAAAAAATTAAGAATGATTTGAATA 1955
OY 366 GACGTCCAGTTCGGTATTAATCTAATTAAC 395
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1954 AATGTCAATTAACAGAAATTAACCTAACAC 1925
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
US-09-129-112-3/c
? Sequence 3, Application US/09129112
? Patent No. 6465716
? GENERAL INFORMATION:
? APPLICANT: Etzler, Marilyn E.
? APPLICANT: Murphy, Judith B.
? APPLICANT: The Regents of the University of California
? TITLE OF INVENTION: A No. 6465716 Factor Binding Protein From Legume Roots
? FILE REFERENCE: 023070-079810US
? CURRENT APPLICATION NUMBER: US/09/129,112
? PRIOR FILING DATE: 1998-08-04
? PRIOR APPLICATION NUMBER: US 08/907,226
? NUMBER OF SEQ ID NOS: 19
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 3
? LENGTH: 6265
? TYPE: DNA
? ORGANISM: Dolichos biflorus
? FEATURE:
? OTHER INFORMATION: genomic sequence of NBP46 (DB46)
? NAME/KEY: exon
? LOCATION: (633)..(944)
? NAME/KEY: Intron
? LOCATION: (945)..(1022)
? NAME/KEY: exon
? LOCATION: (1023)..(1151)
? NAME/KEY: Intron
? LOCATION: (1152)..(1559)
? NAME/KEY: exon
? LOCATION: (1560)..(1616)
? NAME/KEY: Intron
? LOCATION: (1617)..(1697)
? NAME/KEY: exon
? LOCATION: (1698)..(1790)
US-09-129-112-3

```

```

Query Match          4.58; Score 49.2; DB 4; Length 6265;
Best Local Similarity 53.1%; Pred. NO.0.023;
Matches 173; Conservative 0; Mismatches 148; Indels 5; Gaps 3;

OY 15 TAATATTAGTCAATTAACAAATGCATAGAAAGTTCCAAAAAAATTTGTTACAGAAACT 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2936 TATTTATAGAGTGTWTACTATTTCAAATTAACATTAATGCAGTATGAAGTTACTG -AACT 2878
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 75 TCCAAATTTTATTTTATTTATGACAAAGAAATTAACAGATGAAAACATTTTGTGTGGA 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2877 TCTTTTATTTTTTTTTTTTTATTTGAAAAAGTTTGTGAAAAATTTGTCAAAATTTAAATATAGTGA 2818
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 135 ATGCAAGTAGTAATATACATTA -GCCAATTTTAAAAAATTTATTTAGCCTTACGCCG 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2817 ATATTATAAGAAAAATGCAATTAATGAAAACTGAAAAAATTAATTTATTAAGTAGA 2758
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 193 TCAAGTAGTGTTCATGATAGTGTGAATTAATA -ATGCATGCTGCATTCAGAAATGGG 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2757 TCACAAACAAGTTTGAAATTTAATTTAAACATATATTAATTAATTAACCAATTAATTTAT 2698
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 251 ACAACAATGAAACAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGAGTAATG 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2697 CTATTAATTAATAAATTAATTTTAATTAATAATTAATTAATTAATTAATTAATTAATGATCATTATGCTA 2638
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 311 TGTTCGACTATTGAGGGGCAAAA 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2637 TTTTATATTATCAATTAAAGCAGACAA 2612
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chlunis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelisen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

```

Query Match          4.28; Score 46; DB 1; Length 731;
Best Local Similarity 46.38; Pred. No. 0.07;

```

	Matches	151;	Conservative	0;	Mismatches	175;	Indels	0;	Gaps	0;
QY	15	TAATATT	AGTCAT	AAACATG	CAT	AGAAAGTTCC	AAAAAAATTTGGT	TAACAGAACT	74	
Db	22	TTATATAT	TGGTAA	TATGCAT	GAATATTTTGAT	TAGATATGTTTTTTT	TCCTTTT		81	
QY	75	TCCAAAT	TTTTTTTTTTTAT	GTGAC	ACAGAAAT	TAACAGAT	AGAAACTATTTTGGT	GTGGA	134	
Db	82	TTTAT	TTTTTATTTTTTATTTTTTTT	AAAAAT	TTGCAAT	TAATATTTT	TATTTGTA	141		
QY	135	ATGGA	AGTAGTAA	TATACATTTAC	TAGCAAAATTTTAAAAAAT	TATATAGCCTAT	ATACGGC	CTC	194	
Db	142	GGAG	TTTTTATTTGATTTTAAAT	TATATATATTA	TAACAT	ATAGCAACCT	AAAAAT	TGATTTG	201	
QY	195	AAAGTA	TGTTATCT	GTAGGTGTAATTAAT	TAATGCA	TGGTGCAT	CAGATTCAG	AATTTGG	254	
Db	202	ACGGT	ATATGAT	TAAGACAAATTTCT	AAAAAAATTTCCAGT	ATATTTTGGAT	TGGGAAAC	261		
QY	255	CAAT	AAAAACGAAT	TTAAATATTTAC	CTTAAAT	TAATAAATAATTTTG	GTAAATG	CTG	314	
Db	262	CAAC	CAAAAAAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAT	CAAAAAAAAAAAAAAAAA	AAAAAAAAAAT	TAAA	321		
QY	315	TTCT	GACTATTG	AGGGC	AAAAAAAAA	340				
Db	322	ATCA	AAAAAAAAAAGTATTTAA	AGAA	347					

```

RESULT 6
US-08-883-795A-36
// Sequence 36, Application US/08883795A
// Patent No. 5985607
// GENERAL INFORMATION:
// APPLICANT: Delcave, Genevieve
// APPLICANT: Awang, Gregor
// TITLE OF INVENTION: Recombinant DNA Molecules and Expression
// TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
// NUMBER OF SEQUENCES: 39
// CORRESPONDENCE ADDRESS:
// ADDRESSEE: BERESKIN & PARR
// STREET: 40 King Street West
// CITY: Toronto
// STATE: Ontario
// COUNTRY: Canada
// ZIP: M5H 3Y2
// COMPUTER READABLE FORM:
// MEDIUM TYPE: Floppy disk
// COMPUTER: IBM PC compatible
// OPERATING SYSTEM: PC-DOS/MS-DOS
// SOFTWARE: PatentIn Release #1.0, Version #1.25
// CURRENT APPLICATION DATA:
// APPLICATION NUMBER: US/08/883,795A
// FILING DATE: 27-JUN-1997
// CLASSIFICATION: 435
// ATTORNEY/AGENT INFORMATION:
// NAME: Gravelle, Michelle
// REGISTRATION NUMBER: 40,261
// REFERENCE/DOCKET NUMBER: 7841-062
// TELECOMMUNICATION INFORMATION:
// TELEPHONE: (416) 364-7311
// TELEFAX: (416) 361-1398
// INFORMATION FOR SEQ ID NO: 36:
// SEQUENCE CHARACTERISTICS:
// LENGTH: 665 base pairs
// TYPE: nucleic acid
// STRANDEDNESS: single
// TOPOLOGY: linear
// MOLECULE TYPE: cDNA
// ORIGINAL SOURCE:
// ORGANISM: Homo sapiens
// IMMEDIATE SOURCE:
// CLONE: Rh 32
// US-08-883-795A-36

```

Query Match	Similarity	4.28;	Score	45.8;	DB	2;	Length	665;
Best Local	Similarity	47.76;	Pred.	No.	0.076;			
Matches	200;	Conservative	0;	Mismatches	212;	Indels	7;	Gaps
Qy	586	TTTATTAATTTAAAAAACACACAGACTGATCGATGAATAAATTTTCATAAACACATTTTCAGATG	643					
Db	37	TTTTATTAATTTAAAAATTTTAAATTTAAAAATTTATATATTAATTAATTTTAAATTTAAAT	96					
Qy	646	GATTCGTACGATCTCTATAGACTTTTTTTTCTACACAGGTGATGAAGTTATAGT	705					
Db	97	ATTATTAATTTAAATTTTATTAATTTAAAAATTTTATATTAATTTAAATTTTAAATTTAAAT	156					
Qy	706	ACTATTACCGAGACAATGATTATAGATATATACATATTCATATCCATGATATTTTTCGATAT	765					
Db	157	A--TTTATTAATTTAAATTTTATTAATTTAAATTTTAAATTTTAAATTTTATTAATTTAA	214					
Qy	766	AAATAGCGTTAACTATTTCAGCATCGCAGCTTTCGCAACTTTGTTTAAATTTAAG	823					
Db	215	ATATTTAATTAATTAATTTTATTAATTTAAATTAATTTTAAATTAATTTTAAATTTAAAT	274					
Qy	826	AGTTTAATTAATTAATTAAGATTAAAGACATTAAGCAGACAAAGTAATGAAACCGG	885					
Db	275	ATGTTTAATTAATTAATTTATTAATTAATTAATGTTTAATTTACATATTTTAAATTTAA	334					
Qy	886	AGAAACAAAGCCATGAAGCTCATTTGGTTAGCTTAATAGCTTAATPAGAAGATTTTATTTAA	945					
Db	335	ATGTTTAATTAATTT-----ACATATTTTTAATTAATTAAGTTTATATTAATTAATTTTAA	389					
Qy	946	TTTTTAATGACGTGATACATTAATTTATTTTCGACTCTTTAAANCCCCCTACGAA	1004					
Db	330	TTAAATATGTTTATATTCATATTTTAAATTTAAATTAAGTTTATATTAATTTACATATTTTAA	448					

RESULT 7
 US-08-947-823-1
 ; Sequence 1, Application US/08947823
 ; Patent No. 6114605
 ; GENERAL INFORMATION:
 ; APPLICANT: Williamson, Valerie M.
 ; APPLICANT: Kaloshian, Isagouhi
 ; APPLICANT: yaghooobi, Jafar
 ; APPLICANT: Bodeau, John
 ; APPLICANT: Milligan, Stephen
 ; TITLE OF INVENTION: Procedures and Materials for Confering
 ; TITLE OF INVENTION: Pest Resistance in Plants
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentln Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/947,823
 ; FILING DATE: 09-OCT-1997
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US97/18802
 ; FILING DATE: 09-OCT-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/028,191
 ; FILING DATE: 10-OCT-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bastian, Kevin L.
 ; REGISTRATION NUMBER: 34,774
 ; REFERENCE/DOCKET NUMBER: 023070-07021005
 ; TELECOMMUNICATION INFORMATION:


```

NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535
Query Match 4.1% Score 45; DB 4; Length 20674;
Best Local Similarity 45.9%; Pred. No. 0.27;
Matches 153; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
QY 1 AACGACAAATTTATTAAATTTAGTCATPACATAATGCATAGCAACTGCCAAAAAATTTT 60
DB 11391 AATATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11332
QY 61 TGTTAACGAACAACCTCCAAAATTTTTTTTTTTATNGGAACAAGAATAACAGATAGCAAAC 120
DB 11331 AAATTAATTAATTAATTAATTAAGTTAAATTAATTAATTAATTAATTAATTAATTAAT 11272
QY 121 TAATTTGTGGTGGAATGGAAGTAGTAATATACATTACGCAATTTTAAAAAATTAATPA 180
DB 11271 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAAAAAT 11212
QY 181 GCCTATACCGCCTCAAGATATGTATCTAGTAGGTATTAATTAATTAATTAATCATGCGCAT 240
DB 11211 AAAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 11152
QY 241 CAGATTGGGACACAAATGAAAAAGCAATTAATAATTAATTAATTAATTAATTAATTAAT 300
DB 11151 TAAATTTAAATTAATTAATTAATTTTAATTAATTAATTAATTAATTAATTAATTAAT 11092
QY 301 TGAGTAAATGTGTTTCTGACTATTTAGGCGCA 333
DB 11091 TAAATATTCTCTAGCTATTAAAAAAGCA 11059
RESULT 9
US-07-867-106-4/C
Sequence 4, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Ketch L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & NO. 5389526tris
STREET: One Liberty Place 46th Floor
City: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
```



```

?
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/07/867,106
? FILING DATE: 19920625
?
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: AU PJ 7187
? APPLICATION NUMBER: PCV/AN90/00530
? FILING DATE: 02-NOV-1989
?
? ATTORNEY/AGENT INFORMATION:
? NAME: Feeney, Joanne Longo
? REGISTRATION NUMBER: 35,134
? REFERENCE/DOCKET NUMBER: RICE-0002
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215-568-3100
? TELEFAX: 215-568-3439
?
? INFORMATION FOR SEQ ID NO: 4:
?
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3138 base pairs
? TYPE: NUCLEIC ACID
? STRANDEDNESS: single
? TOPOLOGY: circular
?
? MOLECULE TYPE: DNA (genomic)
?
? ANTI-SENSE: NO
?
? US-07-867-106-4

```

Query Match	4.1%	Score 44.8;	DB 1;	Length 3138;
Best Local Similarity	55.9%;	Pred. No. 0.13;		
Matches	85;	Conservative	0;	Mismatches 67;
			Indels	0;
			Gaps	0;

QY	79	AATTTTCTTTTTTATGACACAAGAAATAACGGTAGAAGAATATTTTGTTGTGGCAATGG	138
Db	2047	AATTATTTCTGTTTAAGCAAATAAGAAAAAAAATCTTTTATAGCAATTCGA	1988
QY	139	AAGAGTAATTTACTATTAGCAAAATTTTAAAAATTTATATAGCCTATAGCGGCTCAAG	198
Db	1987	AAATTTTGAATCCCATTT	1928
QY	199	TATGTATCTAGTGC GTATTAATTAATGCA	230
Db	1927	TTTTTTTTTAAITTTGGGTTTTTAAAAATTTTCA	1896

RESULT 10
US-08-487-826B-13/C
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Slim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEO ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826E-13

Query Match	4.18;	Score 44.6;	DB 2;	Length 19124;
Best Local Similarity	47.38;	Pred. No. 0.33;		
Matches 169;	Conservative 0;	Mismatches 184;	Indels 4;	Gaps 1

[illegible]

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINBER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:25:26 ; Search time 83.8565 Seconds

(without alignments)
11172.915 Million cell updates/sec

Title: US-09-502-426a-1_COPY_2102_3202

Perfect score: 1101
Sequence: 1 aactcaacaattataat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.4	4.9	516	10	US-09-960-352-5785 Sequence 5785, Ap
2	53.2	4.8	431	10	US-09-960-352-5558 Sequence 5558, Ap
3	52.2	4.7	377	10	US-09-960-352-7419 Sequence 7419, Ap
4	52.2	4.7	446	10	US-09-960-352-3400 Sequence 3400, Ap
5	49.2	4.5	6265	10	US-09-129-112-3 Sequence 3, Appl1
6	48.6	4.4	53332	9	US-10-224-562-3 Sequence 3, Appl1
7	48.6	4.4	53332	10	US-09-801-861-3 Sequence 3, Appl1
8	48.4	4.4	406	10	US-09-960-352-10265 Sequence 10265, A
9	47.8	4.3	294	10	US-09-960-352-4637 Sequence 4637, Ap
10	47.2	4.3	419	9	US-09-960-352-11234 Sequence 11234, A
11	46.2	4.2	1713	10	US-09-938-842A-4756 Sequence 4756, Ap
12	46	4.2	32463	9	US-09-996-956-5 Sequence 5, Appl1
13	45.8	4.2	428	10	US-09-960-352-573 Sequence 573, App
14	45.6	4.1	1109	9	US-09-938-842A-3333 Sequence 3333, Ap
15	45	4.1	413	10	US-09-960-352-2819 Sequence 2819, Ap
16	44.2	4.0	344	10	US-09-960-352-1036 Sequence 1036, Ap
17	44.2	4.0	4187	9	US-10-072-349-252 Sequence 252, App
18	44.2	4.0	4187	10	US-09-764-855-252 Sequence 252, App
19	44	4.0	376	10	US-09-960-352-5087 Sequence 5087, Ap

C	20	44	4.0	513509	9	US-09-754-853A-4	Sequence 4, Appl1
C	21	44	4.0	640681	10	US-09-790-988-1	Sequence 1, Appl1
C	22	43.8	4.0	4442	10	US-09-960-352-12911	Sequence 12911, A
C	23	43.8	4.0	2000	9	US-09-938-842A-3251	Sequence 3251, Ap
C	24	43.6	4.0	393	10	US-09-960-352-4582	Sequence 4582, Ap
C	25	43.4	3.9	2000	9	US-09-938-842A-4975	Sequence 4975, Ap
C	26	43.2	3.9	416	10	US-09-960-352-4584	Sequence 4584, Ap
C	27	42.8	3.9	640681	10	US-09-790-988-1	Sequence 1, Appl1
C	28	42.6	3.9	214	10	US-09-960-352-7907	Sequence 7907, Ap
C	29	42.6	3.9	2000	9	US-09-938-842A-4090	Sequence 4090, Ap
C	30	42.4	3.9	351	9	US-09-803-719-1436	Sequence 1436, Ap
C	31	42.4	3.9	424	10	US-09-960-352-11218	Sequence 937, App
C	32	42.4	3.9	437	10	US-09-969-373-937	Sequence 10262, A
C	33	42.4	3.9	451	10	US-09-960-352-10262	Sequence 102, App
C	34	42.4	3.9	470	9	US-09-232-880-102	Sequence 102, App
C	35	42.4	3.9	470	9	US-10-012-896-102	Sequence 102, App
C	36	42.4	3.9	470	9	US-09-895-793-102	Sequence 102, App
C	37	42.4	3.9	470	9	US-09-895-814-102	Sequence 102, App
C	38	42.4	3.9	470	10	US-09-759-143-102	Sequence 102, App
C	39	42.4	3.9	470	10	US-09-780-666-102	Sequence 102, App
C	40	42.4	3.9	470	10	US-09-030-606-102	Sequence 102, App
C	41	42.4	3.9	470	10	US-09-822-827-102	Sequence 102, App
C	42	42.4	3.9	470	10	US-09-115-453-102	Sequence 102, App
C	43	42.4	3.9	1045	9	US-09-938-842A-3089	Sequence 3089, Ap
C	44	42.4	3.9	2000	9	US-09-938-842A-4902	Sequence 4902, Ap
C	45	42.4	3.9	335913	9	US-09-754-853A-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-09-960-352-5785/C
; Sequence 5785, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5785
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (76) (90)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1
; US-09-960-352-5785

Query Match 4.9%; Score 54.4; DB 10; Length 516;
Best Local Similarity 48.7%; Pred. No. 0.09; Mismatches 156; Indels 0; Gaps 0;
Matches 148; Conservative 0;

QY	5	TACAAATTATTAATTAGTCATACAAATGCAAGTTCACAAAATTTGTT	64
DB	460	TAAATAAATAATTAATTAATAAATAAATAAATAAATAAATAAATAA	401
QY	65	AACAGAACTCCAAATTTTATTAATGCAAGCAATTAACAGATGAAACTATT	124
DB	400	AAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAA	341
QY	125	TGCTTGTAATGAGTAGTATATACATTAAGCAATTTAAATAATTAATTAAGCCT	184
DB	340	TTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	281

Qy	185	ATACCCCTCAACTAGTTTCTACGAGGCTAATTAATATCAGTCGCGCTCAGA	244
Db	280	ATATPAAAAATPAAATAA	221
Qy	245	ATTGGCAACAATGAAGCGAATTAACTTAACTTTAAATTAATTAATTTTGAG	304
Db	220	AAAAAAAAATTAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAAAAAAAATA	161
Qy	305	TAAA	308
Db	160	TAAA	157

```

RESULT 2
US-09-960-352-5558/c
: Sequence 5558, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengding
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION ANT
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21.10298/C
: CURRENT APPLICATION NUMBER: US/09/960.352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 5558
:   LENGTH: 431
:   TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 24-LIB3057-024-Q1-K1-F7
US-09-960-352-5558

```

Query Match	4.8%;	Score 53.2;	DB 10;	Length 431;
Best Local Similarity	46.1%;	Pred. No. 0.15;		
Matches 178;	Conservative	0;	Mismatches 208;	Indels 0;
			Gaps	0;
QY 16	AATATTGTGCATATACATGCATGAAAGTTCCAAAAAATTTTGTATACAGAAACTT	75		
Db 429	AATAATATAATAAAAATTA	370		
QY 76	CCAAATTTTTTTTTTTTTATGGAACAAGAAATATACAGATAGAAACTATTTTGTGTGGAA	135		
Db 369	ATATAAAATTAATTAATATAAAAAAAAAATATAAAAAAAAAAAAAAAAAAAAAAAAATATAAT	310		
QY 136	TGGAGTAGTATATATACATTAAAGCAAAATTTTAAAAAATATATTAAGCCATACGGGCTCA	195		
Db 309	AATATAAAATATATAAAACAATATAAAAAAAAAAAAAAAAAAATATAATATAAAAAAAAAA	250		
QY 196	AAGTATGTTATCTAGTACGTGTAATTAATATGCAATGGTGGCATTCCAGATTGGGCAAC	255		
Db 249	AATAATATAA	190		
QY 256	AATGAAAAAGCAATTAATAATATTAACTTTAAATATAAATTTTGAATAAATGCTT	315		
Db 189	AAAT	130		
QY 316	TCTGACTATTGAGGGGCAAAAAAAAAAGACATGCCAAAAGCTACGGGTTTGACTGTCCAG	375		
Db 129	AAAAAAAAAATAAAAAAAAAAATAAAAAAAAAAATAAATAAATAAATTTATAGTAAAC	70		
QY 376	TTTCGTAATATCTATAATACCTGTC	401		
Db 69	CAAAATTATAAATAAACCAACAGCTC	44		

RESULT 3
US-09-960-352-7419/c
; Sequence 7419, Application US/09960352
; Patent No. US20020137139a1
; GENERAL INFORMATION:

```

: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengbing
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960,352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 7419
:
: LENGTH: 377
:
: TYPE: DNA
:
: ORGANISM: Bos taurus
:
: OTHER INFORMATION: Clone ID: 32-LIB3057-025-Q1-K1-H11
US-09-960-352-7419

```

	Query Match Similarity	4.7%;	Score 52.2;	DB 10;	Length 377;	
	Best Local Similarity	49.5%;	Pred. No. 0.22;			
	Matches 161;	Conservative	0;	Mismatches 163;	Indels 1;	Gaps 1
OY	570	TATTAGCAGAATGGGTTTATTAATTTAAAACTCACACCTTGATGCATATAAATTCAT	629			
Db	351	TTTTGAATGTACGCTTTTGTAATTTAGATTTTTTTTATCGTGAGACTTTTTTTTTT	292			
OY	630	AAACCTTTTACGANNGATCGTACCAGTCATCAATGACTTTTTTTTTTTCACACGGT	689			
Db	291	TTAGAAATTTTTTAGGTATTTTTTTTTTGTATAATAGTTAATATTAATTTATTTTATA	232			
OY	690	GGATGAAGTATTTACAGCATAGCCAGACAATTTGATTATAGATATATTCATTAATCC	749			
Db	231	TTTTGATTTTACATATTTTATTTAAAATTTTAGCTTG-TTTTAAATTAATTTATTAATGT	173			
OY	750	ATGATATTTATGATATTAATAGCTGTAAACATTTTCAGACATCGACCTTTCGAACCT	809			
Db	172	TGGATTTTTTTTATTTTCANTATTTTAAATTTTATTTTTTTTTTTTTTTTAAATTATA	113			
OY	810	TTTGTTTTPAATTTAAGAGTTTATTAATTAATAAGTATTTAAAGAGCATACGAGCCACA	869			
Db	112	TTATTTTPTTTTATTTTAAATTAATAATTAATAATATATAATATAAANAACAAACA	53			
OY	870	AAAGTAATGAACACGAGAAACAA	894			
Db	52	AAAAAAAAAAAAAAAAAAAAA	28			

RESULT 4
 US-09-960-352-3400/C
 Sequence 3400, Application US/09960352
 Patent No. US2002013719A1
 GENERAL INFORMATION:
 APPLICANT: Warren, Wesley C.
 APPLICANT: Tao, Mengbing
 APPLICANT: Byatt, John C.
 APPLICANT: Mathialagan, Nagappan
 TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 FILE REFERENCE: 16511.006/37-21(1028)C
 CURRENT APPLICATION NUMBER: US/09/960.352
 CURRENT FILING DATE: 2001-09-24
 NUMBER OF SEQ ID NOS: 15112
 SEQ ID NO 3400
 LENGTH: 446
 TYPE: DNA
 ORGANISM: Bos taurus
 OTHER INFORMATION: Clone ID: 15-LIB3058-052-01-K1-D11
 US-09-960-352-3400

Query Match	4.7%;	Score 52.2;	DB 10;	Length 446;
Best Local Similarity	46.7%;	Pred. No. 0.23;		
Matches 165;	Conservative	0;	Mismatches 188;	Indels 0;
				Gaps 0;
QY	1	AATCTACCAATTTATTTAATTTTATGTCATTAACAAATGCATAGAAAGTTCCAAAAAATTTT	60	

Db 427 AAAAAAAAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATA 368
 Oy 61 TGTAAACAAACTCCAAATTTTTTTTTTATGGAACAGAAATTAAGATGAAC 120
 Db 367 AATATATATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATA 308
 Oy 121 TATTTTGTGTGATGAGAGTAATATACATTAAACCAATTTTAAAAATATATA 180
 Db 307 TAAAAAAAAAATATATAATTAATTAATAAAAAAAAAAAAAAAAAATA 248
 Oy 181 GCCTTACGGCGCTCAAGTATGTTATGTAGTGGTGAATTAATATGATGGTGATT 240
 Db 247 AAAAAAAAAAAAAAAAAATATATAAAAAAAAAAAAAAAAAAAAAAAAA 188
 Oy 241 CAGATTGGAGCAACATGAAACCGAATTAAATATTAACCTTAAATTAATATAAT 300
 Db 187 AAA 128
 Oy 301 TGAATTAATGTGTTTCTGACTATTGAGGGCAAAAAAAAAAGACATGCCAAA 353
 Db 127 AAA 75

RESULT 5
 US-09-129-112-3/c
 ; Sequence 3, Application US/09129112
 ; Patent No. US2002001995A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Etzler, Marilyn E.
 ; APPLICANT: Murphy, Judith B.
 ; APPLICANT: The Regents of the University of California
 ; TITLE OF INVENTION: A No. US2002001995A1 Factor Binding Protein From Legume Roots
 ; FILE REFERENCE: 023070-079810US
 ; CURRENT APPLICATION NUMBER: US/09/129,112
 ; PRIOR FILING DATE: 1998-08-04
 ; PRIOR APPLICATION NUMBER: US 08/907,226
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3

```

1 ORGANISM: Dolichos biflorus
2
3 FEATURE:
4
5 OTHER INFORMATION: genomic sequence of NBPA6 (DB46)
6
7 NAME/KEY: exon
8 LOCATION: (633)..(944)
9 NAME/KEY: intron
10 LOCATION: (945)..(1022)
11 NAME/KEY: exon
12 LOCATION: (1023)..(1151)
13 NAME/KEY: intron
14 LOCATION: (1152)..(1559)
15 NAME/KEY: exon
16 LOCATION: (1560)..(1616)
17 NAME/KEY: intron
18 LOCATION: (1617)..(1697)
19 NAME/KEY: exon
20 LOCATION: (1698)..(1790)
21
22 US-09-129-112-3

```

Query Match	4.58;	Score 49.2;	DB 10;	Length 6265;
Best Local Similarity	53.18;	Pred. No. 2.4;		
Matches 1/3;	Conservative	0;	Mismatches 148;	Indels 5;
			Gaps	3
QY	15	TAAATATTAGTCAATATACATGCATAGAAAGTCCAAAAAATTTTGTGTAACAGAAACT	74	
Db	2936	TATTTATATAGTGTCTTACTATTCAATATACATATATGCAGTATGAAGTTACTGTG-AACT	2878	
QY	75	TCCAAATTTTTTTTTTTTATGACACAGAAATTAACAGATAGAAAACATTTTCTGTGGA	134	
Db	2877	TCATTTTTTTTTTTTTTTATGAAAATGTTTGAAAATATGTCAANATTAATATAGTGA	2818	

QY 135 ATGGAAGTAGTAATATACATTA--GCAATTTTAAAAATTATATAGCTATACGCC 192
Db 2817 ATATTATAAGAAAAATGCATATTATGAAAACGAAAAAATAATTAATTATTAAGTAA 2758
QY 193 TCAAGATGTGTATCTAGTAGAGGTAAATTA--ATGCATGTGGCATTCAAGATTGGG 250
Db 2757 TCACACAAAGTTTGAAATTATTTAAACTATATTATAAATATTACCAATACATTTAT 2698
QY 251 ACAACATGAAAAAGCAATTAATATTACTTTAAATTAATTAATTAATTAAGTAATG 310
Db 2657 CTATAAAATTAATTAATTAATTTTAAATATAATTAATTAATTAATTAATTAATTA 2658
QY 311 TGTTCCTGACTTTGAGGGCAAA 336
Db 2637 TTTTATTTATCAATTAACGACACA 2612

```

RESULT 6
US-10-224-562-3/C
; Sequence 3 Application US/10224562
; Publication No US2003022299A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TIME OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TIME OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001098DIV
; CURRENT APPLICATION NUMBER: US/10/224,562
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 53352
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-562-3

```

[illegible]

RESULT 7
 US-09-801-861-3/c
 ; Sequence 3, Application US/09801861
 ; Patent No. US20020119544A1
 ; GENERAL INFORMATION:
 ; APPLICANT: YAN, Chunhua et al.
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
 ; FILE REFERENCE: C1001098

Query Match	4.4%;	Score 48.4;	DB 10;	Length 406;
Best Local Similarity	50.0%;	Pred. No. 1.2;		
Matches 121;	Conservative 0;	Mismatches 121;	Indels 0;	Gaps 0;
QY 57	ATTTTGTTAACAGAAACTTCCAAATTTTTTTTTTTTATGGAACAAGAAATACAGATGCA	116		
DB 275	ATTTTATTTATTTTAAATATGTTGTTTCATTTTAACTTTATTTATTTATATATTTATTCCTTAA	216		
QY 117	AAACATTTTGTGTGCAATGGAAGTACTAATATTCATTAAAGCAAAATTTTAAAAATATAT	176		
DB 215	AAAAATAGATTTATTTAAATATATTTTAAACAAAAATTTATAAAAAAATATATATTAATTAATTA	156		
QY 177	ATAAGCCTATACGGCGCTCAAGATATGTATCTAGTAGTGATTAATTAATATATGATCGATGGTC	236		
DB 155	AAAAATATAAAAATATATAAAAATTAATAAAAATTAATAATTAATTAATTAATTAATTAAT	96		

```

RESULT 10
US-09-960-352-11234/c
Sequence 11234, Application US/09960352
Patent No US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengling
APPLICANT: Byatt, John C.
APPLICANT: Mathalagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(102981C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 11234
LENGTH: 419
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 48-LIB3058-052-Q1-K1-D8
US-09-960-352-11234

```


Query Match 4.3%; Score 47; DB 10; Length 419;
Best Local Similarity 46.1%; Pred. No. 2.3;
Matches 158; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 11 TTATTAATTTTGAATACATGATGAAAGTCCAAAAATTTTGTACAGA 70
DB 413 TTATTAATTTTGAATACATGATGAAAGTCCAAAAATTTTGTACAGA 70
QY 71 AACTCCAAATTTTGTATGCAACAGAAATTAACAGATGAAACTATTTGTC 130
DB 353 ATTTTGTAAATTTTGTATGCAACAGAAATTAACAGATGAAACTATTTGTC 130
QY 131 TGAATGGAAGTAAATATACATTAAGCAATTTTAAATTAATTAAGCTATACGC 190
DB 293 TTAATAATTTTGAATACATGATGAAAGTCCAAAAATTTTGTACAGA 70
QY 191 GCTCAAGTGTATCTAGTGTATGATTAATGATGCTGCTGATGCAATGCG 230
DB 233 TTAATTAATTTTGAATACATGATGAAAGTCCAAAAATTTTGTACAGA 70
QY 251 ACAACAATGAAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 174
DB 173 AAA 114
QY 311 TGTCTCTGCTATGAGGGCAAAAAAAAAAGACATGCGCAAA 353
DB 113 AAA 71

RESULT 11

US-09-938-842A-4756
; Sequence 4756, Application US/0993842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krieps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938, 842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227, 866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264, 647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300, 111
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4756
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4756

Query Match 4.2%; Score 46.2; DB 9; Length 1713;

Best Local Similarity 48.3%; Pred. No. 5.7; Mismatches 138; Indels 0; Gaps 0;

QY 51 AAAAAAAAAATTTGTAACAGAACTCCAAATTTTGTATGCAACAGAAATTAACA 110
DB 874 ATATTAATTTTGAATACATGATGAAAGTCCAAAAATTTTGTACAGA 70
QY 111 GATGAAATTAATTTGTAACAGAACTCCAAATTTTGTATGCAACAGAAATTAACA 170
DB 934 CATATTAATTTGTAACAGAACTCCAAATTTTGTATGCAACAGAAATTAACA 170
QY 171 AATTAATTAACCTTAATGCGCTCAAGTATGTTATCTAGTGTATGCAATTAATTAATGA 230
DB 994 AATTAATTAATTAATTAATTTGTTTATTAATTAATTAATTAATTAATTAATGA 1053

QY 231 TGCTGCATTCAGATTTGGGACAAATGAAACGAAATTAATTAATTAATTAATTA 290
DB 1054 GCTTAATTAATTAAGGAGATGATGCAAGATTAATTAATTAATTAATTAATTA 1113
QY 291 AATTAATTAATTTGATTAATGCTTTTC 317
DB 1114 TATGAAGACTCATCACTTTGATTC 1140

RESULT 12

US-09-966-956-5/C
; Sequence 5, Application US/0996956
; Patent No. US20020155463A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Prostate Polynucleotides and Uses
; FILE REFERENCE: 9U 301 R1
; CURRENT APPLICATION NUMBER: US/09/966, 956
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250, 354
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 32463
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-956-5

Query Match 4.2%; Score 46; DB 9; Length 32463;
Best Local Similarity 47.0%; Pred. No. 19;
Matches 142; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 39 TAGAAGTTCACAAATTTTGTATGCAACAACTCCAAATTTTGTATGCAACAACTCCAA 98
DB 4922 TATTAAGTATTAATTTTGTATGCAACAACTCCAAATTTTGTATGCAACAACTCCAA 4863
QY 99 CAGAAATTAACAGATTAACAACTTTTGTATGCAACAACTTTTGTATGCAACAACTTTTGT 158
DB 4862 ACACAAAGAAATTAACAGATTAACAACTTTTGTATGCAACAACTTTTGTATGCAACAACTTT 4803
QY 159 CAAATTTTAAATTTTATTAAGCTTAACAGCTTAACAGCTTAACAGCTTAACAGCTTA 218
DB 4802 AATTTTGAAGTGTGCTTTGGAATTCAGTCAAGTGTGGAATTCAGTCAAGTGTGGAATTC 4743
QY 219 ATTAATTAAGTGTGCTTTGGAATTCAGTCAAGTGTGGAATTCAGTCAAGTGTGGAATTC 278
DB 4742 TCCAGTTTACTGATTAATTTTGTATGCAACAACTTTTGTATGCAACAACTTTTGTATGCA 4683
QY 279 AACTTAATTAATTAATTAATTTTGTATGCAACAACTTTTGTATGCAACAACTTTTGTATG 338
DB 4682 TATGATTTGATTAATTAATTTTGTATGCAACAACTTTTGTATGCAACAACTTTTGTATG 4623
QY 339 AA 340
DB 4622 GA 4621

RESULT 13

US-09-960-352-573/C
; Sequence 573, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengling
; APPLICANT: Byate, John C.
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960, 352
; NUMBER OF SEQ ID NOS: 15112

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:08:52 ; Search time 1436.66 Seconds

(without alignments)
12411.569 Million cell updates/sec

Title: US-09-502-426a-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatctacaattatataat.....gagagagagaactagcttc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 / Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estipl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	13.4	148	BH811044	BH811044 SALK_0571
2	145.6	13.2	473	AL768954	AL768954 Arabidops
3	69.6	6.3	1101	CNS0039G	AL063921 Drosophill
4	69.4	6.3	1201	CNS0167M	AL106336 Drosophill
5	66.8	6.1	928	CNS000DKY	AL071865 Drosophill
6	66.4	6.0	1201	CNS0165X	AL106335 Drosophill

7	65.8	6.0	1101	17	CNS0021J	AL061936 Drosophill
8	65.2	5.9	1101	17	CNS001FT	AL078714 Drosophill
9	64.6	5.9	1101	17	CNS008EVL	AL069706 Drosophill
10	63.6	5.8	1146	17	CNS021G2	AL176843 Tetradon
11	63.4	5.8	1101	17	CNS0021J	AL061936 Drosophill
12	63.2	5.7	1101	17	CNS0080L	AL057419 Drosophill
13	62.8	5.7	1009	17	CNS00587	AL057153 Drosophill
14	62.6	5.7	1101	17	CNS008EVL	AL069706 Drosophill
15	62.4	5.7	855	17	CNS04P5D	AL069706 Drosophill
16	62	5.6	1029	17	CNS014GM	AL069706 Drosophill
17	61.8	5.6	1101	17	CNS0033D	AL174271 Tetradon
18	61.6	5.6	609	17	CNS023K2	AL064091 Drosophill
19	61.6	5.6	1101	17	CNS016L1	AL182171 Tetradon
20	61.6	5.6	1190	17	CNS020N7	AL106896 Drosophill
21	61.6	5.6	1200	17	CNS016CO	AL106896 Drosophill
22	61.4	5.6	1092	17	CNS020K7	AL175656 Tetradon
23	61	5.5	910	17	CNS0188P	AL142826 Anopheles
24	61	5.5	1085	17	AG075009	AG075009 Pan trogl
25	60.8	5.5	1101	17	CNS0080L	AL057419 Drosophill
26	60.8	5.5	1101	17	CNS001FB	AL060732 Drosophill
27	60.6	5.5	878	17	CNS0187R	AL108893 Drosophill
28	60.6	5.5	1101	17	CNS0039G	AL063921 Drosophill
29	60.2	5.5	1201	17	CNS016C2	AL106556 Drosophill
30	60	5.4	945	17	CNS04D0K	AL285149 Tetradon
31	60	5.4	1225	17	CNS0161D	AL106171 Drosophill
32	59.8	5.4	964	17	CNS0788R	AL441457 T7 end of
33	59.8	5.4	966	17	CNS0057C	AL061991 Drosophill
34	59.6	5.4	1092	17	CNS020K7	AL175656 Tetradon
35	59.6	5.4	1368	10	BE420618	BE420618 HMM000.60
36	59.2	5.4	524	17	CNS01090	AL167541 Tetradon
37	59.2	5.4	843	17	CNS00CS1	AL059666 Drosophill
38	59	5.4	987	17	CNS014PQ	AL104456 Drosophill
39	59	5.4	984	17	CNS04NDJ	AL298972 Tetradon
40	58.8	5.3	1101	17	CNS008X3	AL052544 Drosophill
41	58.6	5.3	1001	17	CNS0135H	AL105023 Drosophill
42	58.6	5.3	1101	17	CNS00807	AL069440 Drosophill
43	58.4	5.3	676	17	CNS003XU	AL065304 Drosophill
44	58.4	5.3	843	17	CNS00CS1	AL059666 Drosophill
45	58.4	5.3	996	17	CNS008FH	AL071063 Drosophill

ALIGNMENTS

RESULT 1
BH811044/c 148 bp DNA linear GSS 02-MAY-2002
SALK_057128 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_057128, DNA sequence.

ACCESSION
BH811044
VERSION
KEYWORDS
SOURCE
ORGANISM

GSS.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadriab,
'C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Predits,I., Shinn,P.,
'Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the

TITLE
JOURNAL
COMMENT

Unpublished (2001)
Arabidopsis Genome
Contact: Joseph R. Ecker
The Salk Institute for Biological Studies (SIGMA)

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.
 FEATURES Location/Qualifiers
 source 1.148
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_057128"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 46 a 28 c 14 g 60 t
 ORIGIN

Query Match 13.4%; Score 148; DB 17; Length 148;
 Best Local Similarity 100.0%; Pred. No. 6.6e-18;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 ATGGAACAAGAAATACAGATAGAAACTATTTGTTGGATGAGAGTAGATATATAC 152
 DB 148 ATGGAACAAGAAATACAGATAGAAACTATTTGTTGGATGAGAGTAGATATATAC 89
 OY 153 ATTAAGCAAAATTTTAAAAATATATACCTATACGGCTCAAGATATATCTAGTA 212
 DB 88 ATTAAGCAAAATTTTAAAAATATATACCTATACGGCTCAAGATATATCTAGTA 29
 OY 213 GGTGTAATTAATGATGATGTCGATTT 240
 DB 28 GGTGTAATTAATGATGATGTCGATTT 1

RESULT 2
 AL768954 473 bp DNA linear GSS 18-JUN-2002
 LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-082A08-011867,
 DEFINITION genomic survey sequence.
 ACCESSION AL768954
 VERSION AL768954.1 GI:21522073
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1
 Stelzhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Siedler, H.
 and Weissshaar, B.
 A pipeline for automated high-throughput generation of ESTs
 (flanking sequence tags) from Arabidopsis thaliana T-DNA
 transformed lines
 unpublished
 2
 Rosso, M., Stelzhov, N., Li, Y., Reiss, B., Dekker, K. and Weissshaar, B.
 A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
 for flanking sequence tag based reverse genetics
 unpublished
 3 (bases 1 to 473)
 Li, Y., Rosso, M., Stelzhov, N. and Weissshaar, B.
 Direct Submission
 Submitted (17-JUN-2002) Weissshaar B., Max-Planck-Institut fuer
 Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion within the locus defined by clone T345. The
 sequences are generated at the MPI for Plant Breeding Research in
 the context of the GABI-Kat project. GABI-Kat is part of the German
 Plant Genomics program designated 'GABI'. Information on line
 availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.
 Location/Qualifiers
 1. 473

/organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="GK-082A08-011867"
 /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 168 a 58 c 70 g 146 t 31 others
 ORIGIN

Query Match 13.2%; Score 145.6; DB 17; Length 473;
 Best Local Similarity 87.0%; Pred. No. 1.3e-17;
 Matches 160; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 1 AATCTACAATTTATATATTATGTCATTAACATGATGAAAGTCCAAAAATTTT 60
 DB 290 AATCTACAATTTATATATTATGTCATTAACATGATGAAAGTCCAAAAATTTT 349
 OY 61 TGTTAACAGAACTTCCAAATTTTTTTTTTTATGGAACAGAAATACAGATGAAAAC 120
 DB 350 TGTTAACAGAACTTCCAAATTTTTTTTTTTGCGGCGCAAAATATCCGGATGGAACC 409
 OY 121 TATTTTGTGGAAGGAGTAGTATATACATTAAGCAAAATTTTAAAAATATATTA 180
 DB 410 TTTTGTGCGGCGGAGGAGGAGTATATATATCTTAACCAAAATTTTAAAAATTTTAA 469
 OY 181 GCCT 184
 DB 470 CCTT 473

RESULT 3
 CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TETS end of BAC #
 DEFINITION BACR08K10 of RCI1-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL063921
 VERSION AL063921.1 GI:4941778
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqif@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutyo Osoegawa and
 Aaron Mamoser in Pletier de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RCI1-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 p1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

SOURCE ORGANISM
<i>Drosophila melanogaster</i> .
<i>Drosophila melanogaster</i>
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

REFERENCE
AUTHORS
TITLE
JOURNAL

Ephydroidae; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mammose in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

COMMENT

FEATURES
SOURCE

1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_1lb="RPCI-98"
/note="end : 17"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 5.9%; Score 64.6; DB 17; Length 1101;
Best Local Similarity 34.3%; Pred. No. 0.014;
Matches 139; Conservative 71; Mismatches 195; Indels 0; Gaps 0;

OY 586 TTTATTATTTAAACACACACACACACATTAATTAATTAACCTTTTACGANG 645
DB 915 TTTATTTTATTTATTTATTTTAAATTTTATTTTATTTTATTTATTTATTTA 856
OY 646 GATTGTCAGATCTATCTATGACTTTTCTTACACGCGTGGATGAAGTTACT 705
DB 855 TTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 796
OY 706 ACTATGACGACGACATGATTTATGATATATCCATTAATCATGATATTTAGAT 765
DB 795 ATTATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 736
OY 766 AATAGCTTTAACTATTTGACGACGCGCTTCTGCACTTTTCTTTATTTAG 825
DB 735 ATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 676
OY 826 AGTTATATTTAAATTAAGTATTAAGGACATACGACGACGACGACGACGACG 885
DB 675 TTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 616
OY 886 AGAATTAATTAAGGACGATGATTTAGTTAAGTTAATTAAGGATTTATTA 945
DB 615 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 556
OY 946 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 990
DB 555 TTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 511

RESULT 10
LOCUS CNS021G2 1146 bp DNA linear GSS 12-MAY-2000
DEFINITION Tetradon nigroviridis genome survey sequence 17 end of clone
225004 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL176843
VERSION AL176843.1 GI:7814900

KEYWORDS
SOURCE

GSS; genome survey sequence.
Tetradon nigroviridis.

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Perciformes; Tetracnontiformes;
Tetracnontidae; Tetracnont.

REFERENCE
AUTHORS

1 (bases 1 to 1146)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Filzmes, C., Wnicker, P., Brotlier, P., Quetier, F.,
Sautin, W. and Weissenbach, J.

TITLE

Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence

JOURNAL
AUTHORS

Unpublished
2 (bases 1 to 1146)
Roest-Crollius, H., Jallion, O., Dasilva, C., Filzmes, C.,
Bouneau, L., Billault, A., Quetier, F., Sautin, W., Bernot, A. and
Weissenbach, J.

TITLE

Characterization and repeat analysis of the compact genome of the
freshwater puterfish Tetradon nigroviridis

JOURNAL
AUTHORS

Unpublished
3 (bases 1 to 1146)
Genoscope.

REFERENCE
AUTHORS

Submitted (12-Apr-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetradon>.

FEATURES
SOURCE

1. 1146
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="225004"
/clone_1lb="G"
/note="Genoscope sequence ID : C0AG225BH02LPI-end : 17"

BASE COUNT 393 a 60 c 73 g 274 t 346 others
ORIGIN

Query Match 5.8%; Score 63.6; DB 17; Length 1146;
Best Local Similarity 31.8%; Pred. No. 0.021;
Matches 150; Conservative 98; Mismatches 222; Indels 1; Gaps 1;

OY 534 ACTTCCTGGATTCCTGCGACGATTTATCAAAAATTTATGACGACGATGCTTTA 593
DB 591 ATTTATTTTATTTTAAKRRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKR 650
OY 594 TTTAAATCTCACACTGATGATCAATTAATTTCAATTAACCTTTTACGATTC 653
DB 651 KTTTAAKRRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKR 709
OY 654 CGATCTCTAATGACTTTTCTTACACGCGTGGATGAAGTTATGATTTAG 713
DB 710 AGKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRR 769
OY 714 CCAGACAAATGATTAATGATATATCATTAATCATGATATTTATTAATAGCT 773
DB 770 KRRKKRRAAARAAKRRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKK 829
OY 774 GTTAAATCTTCCAGATGCGACCTTTCGCACTTTGCTTTTATTTAAGCTTA 833
DB 830 ATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 889
OY 834 AATTAATGATTTAAAGAGATACGACGACGACGACGACGACGACGACGAC 893
DB 890 TTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 949
OY 894 AAGCATGACGATGCTGATTTAGCTTAATTAAGATTTATTAAGATTTAAG 953
DB 950 AAAATTTTTRRDKRTTTTTRTTRTTRTTRTTRTTRTTRTTRTTRTTRT 1009
OY 954 ACGATGATTAACATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1004

DB	1010	ATAAATAATAAAAATTGTTTGTGTTTGGTTTAAANNAAMTTTGAWMAA	1060
RESULT 11	CNS0021J	1101 bp	DNA linear GSS 03-JUN-1999
LOCUS	CNS0021J/c		
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL061936		
VERSION	AL061936.1	GI:4940214	GSS.
SOURCE	Drosophila melanogaster. Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
ORGANISM	1 (bases 1 to 1101)		
REFERENCE	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
FEATURES	source	location/Qualifiers	
	1..1101	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"		
	/clone="BACR05N11"		
	/clone_1lb="RPCI-98"		
	/note="end : TET3"		
BASE COUNT	631 a	7 c	28 g 289 t 146 others
ORIGIN			
	Query Match	5.8%	Score 63.4; DB 17; Length 1101;
	Best Local Similarity	49.2%; Pred. No. 0.023;	
	Matches	88; Conservative 25; Mismatches 66; Indels 0; Gaps 0;	
OY	2 ATCTACAATTTTAAATTTAGTCATATACATGCATAGAAGTCCAAAAAATTTT	61	
DB	742 WTTTTAAAAMTAATTAATTTTAAAMAAAAAAAMTAAAAAATTTTAAATTTAAATTTT	683	
OY	62 GTTACAGAACTCCCAATTTTGTGCAACAGAAATATACAGATAGAAACT	121	
DB	662 ATTAAATAAAMAATWTATTTTATTAATTAATTAATTAATTAATTAATTAATTTTAAATTT	623	
OY	122 ATTTTGTTGGGAATGAAGTAGTATATACATTAAGCAAAATTTTAAAAAATTTATATATA	180	
DB	622 TTTTWWTTTAAATTAATTAATTAATTAATTTTATTAATTTTAAATTTTAAATTTTAAATTA	564	
RESULT 12	CNS00BOI/c	1101 bp	DNA linear GSS 04-JUN-1999
LOCUS	CNS00BOI		
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR2AD09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL057419		
VERSION	AL057419.1	GI:4937885	

[illegible]

RESULT 13
CNS00587/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CNS00587 1009 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR1N01 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL057153.1 GI:4932541
GSS.
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1009)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammose in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1..1009
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR1N01"
/clone_1lb="RPCI-98"
/note="end : TET3"

BASE COUNT 263 a 116 c 96 g 477 t 57 others
ORIGIN

Query Match 5.7%; Score 62.8; DB 17; Length 1009;
Best Local Similarity 46.5%; Pred. No. 0.031;
Matches 148; Conservative 10; Mismatches 160; Indels 0; Gaps 0;

QY 1 AATTCACAAATTTTATTTTATGTCATATGCAATGAGAAAGTCCAAAAAATTTT 60
DB 1006 AATATWAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 947
QY 61 TGTAAACAGAACTTCAATTTTATTTTATGAGACAGAAATATACAGATAGAAAC 120
DB 946 NATAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 887
QY 121 TATTTGTTGTAATGAGATGATATATACATTAAGCAATTTTAAATTTATATA 180
DB 886 AATATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 827
QY 181 GCTTATAGCGCTCAAGATATGTTATGATAGGTATTAATTAATGATGGTGGATT 240
DB 826 TAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 767
QY 241 CAGAAATTTGAGACAAATGAAAGCAATTAATTTTAAATTTTAAATTTTAA 300
DB 766 AATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 707
QY 301 TGAGTAATGTGTTTCT 318
DB 706 TWTATWAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 689

RESULT 14
CNS00587/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CNS00587 1101 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL069706.1 GI:4949849
GSS.
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammose in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_1lb="RPCI-98"
/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 5.7%; Score 62.6; DB 17; Length 1101;
Best Local Similarity 35.1%; Pred. No. 0.032;
Matches 140; Conservative 65; Mismatches 194; Indels 0; Gaps 0;

QY 8 AATTTATTAATTTTATGTCATATGCAATGCAATGCAATGCAATGCAATGCAAT 67
DB 587 AATWTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 646
QY 68 AGAACTTCAATTTTATTTTATGAGACAGAAATATACAGATGAAATTTTGT 127
DB 647 TATTTTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 706
QY 128 TTGTGAATGAGATGATATATACATTAAGCAATTTTAAATTTTAAATTTTAA 187
DB 707 ATTAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 766
QY 188 CGGCTCAAGATGTTATCTAGTATGATTAATTAATGATGCTGATTCAGAAAT 247
DB 767 TTTTATWAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 826
QY 248 GGGACAAATGAAAGCAATTAATTTTAAATTTTAAATTTTAAATTTTAA 307
DB 827 WATATWATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 886
QY 308 ATGTGTTTCTGACTATTTGAGGGCAAAAAAGACATCCCAAGTCTACGGTTTGA 367

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:06:10 ; Search time 31531.6 Seconds

(without alignments)
2955.362 Million cell updates/sec

Title: US-09-502-426a-1_COPY_1_3202

Perfect score: 3202
Sequence: 1 atgtgggtattatattgttgc.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_hg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pln:*
35: em_hg_rod:*
36: em_hg_mam:*
37: em_hg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	3134.2	97.9	84196	8	ATT3A5	AL132979 Arabidops
C 2	1132	35.4	4918	2	AF044216	AF044216 Arabidops
C 3	176.8	5.5	191	11	AL844732	AL844732 Arabidops
C 4	133.2	4.2	104992	2	AC005504	AC005504 Plasmodu
C 5	133.2	4.2	169586	2	AC004157	AC004157 Plasmodu
C 6	130.8	4.1	104992	2	AC005504	AC005504 Plasmodu
C 7	130.8	4.1	169546	2	AC004157	AC004157 Plasmodu
C 8	129.4	4.0	14867	3	AE001398	AE001398 Plasmodu
C 9	128.6	4.0	172816	9	AC093899	AC093899 Homo sapl
C 10	127.8	4.0	86827	3	PFMAL385	AL034556 Plasmodu
C 11	126.4	3.9	133877	2	AC120883	AC120883 Homo sapl
C 12	125.4	3.9	86827	3	PFMAL385	AL034556 Plasmodu
C 13	124.6	3.9	205130	2	AC105425	AC105425 Homo sapl
C 14	124.6	3.9	4601	3	DM011584	U1584 Drosophila
C 15	124.4	3.9	19517	3	DM037541	U37541 Drosophila
C 16	123.4	3.9	67970	3	PFMAL183	AL031746 Plasmodu
C 17	122.6	3.8	4601	3	DM011584	U1584 Drosophila
C 18	122.6	3.8	19517	3	DM037541	U37541 Drosophila
C 19	121.2	3.8	177293	2	AC069525	AC069525 Homo sapl
C 20	121.2	3.8	172816	9	AC093899	AC093899 Homo sapl
C 21	121	3.8	178670	9	AC104073	AC104073 Homo sapl
C 22	118	3.7	67970	3	PFMAL183	AL031746 Plasmodu
C 23	118	3.7	84196	8	ATT3A5	AL132979 Arabidops
C 24	116.4	3.6	72243	9	AL731858	AL731858 Human DNA
C 25	115.6	3.6	205130	2	AC105425	AC105425 Homo sapl
C 26	115.2	3.6	133877	2	AC120883	AC120883 Homo sapl
C 27	114.4	3.6	178670	9	AC104073	AC104073 Homo sapl
C 28	113	3.5	14867	3	AE001398	AE001398 Plasmodu
C 29	112.6	3.5	137889	9	AC073269	AC073269 Homo sapl
C 30	112.2	3.5	206059	2	AC127383	AC127383 Homo sapl
C 31	111.6	3.5	104853	9	AC117444	AC117444 Homo sapl
C 32	111.4	3.5	111861	9	AC069435	AC069435 Homo sapl
C 33	110.2	3.4	15635	3	AB083339	AB083339 Bombyx
C 34	110.2	3.4	15635	3	AB070264	AB070264 Bombyx mo
C 35	109.6	3.4	159618	9	AC096750	AC096750 Homo sapl
C 36	107.8	3.4	152209	9	HS1108D11	AL034419 Human DNA
C 37	107	3.3	105238	9	AC011458	AC011458 Human DNA
C 38	106.6	3.3	106958	9	AL807813	AL807813 Human DNA
C 39	106.4	3.3	155106	9	AC104069	AC104069 Homo sapl
C 40	106.2	3.3	181792	9	AC098822	AC098822 Homo sapl
C 41	105.6	3.3	176174	9	AC007483	AC007483 Homo sapl
C 42	105.6	3.3	177293	2	AC069525	AC069525 Homo sapl
C 43	105.2	3.3	206059	2	AC127383	AC127383 Homo sapl
C 44	104.8	3.3	114276	9	AC011355	AC011355 Homo sapl
C 45	104.6	3.3	127439	9	AC126178	AC126178 Homo sapl

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
ATT3A5	ATT3A5	Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.	AL132979	AL132979.2	GI:6782244	Arabidopsis thaliana.	1 (bases 1 to 84196)	Bloeker, H., Mewes, H.W., Lemcke, K., Meyer, K.F.X., Quetler, F. and Salanoubat, M.
ATT3A5	ATT3A5	84196 bp DNA linear	PLN 26-JAN-2000					
DEFINITION	Arabidopsis thaliana	Arabidopsis thaliana						
VERSION	AL132979	AL132979						
KEYWORDS	Arabidopsis thaliana	Arabidopsis thaliana						
SOURCE	Arabidopsis thaliana	Arabidopsis thaliana						
ORGANISM	Arabidopsis thaliana	Arabidopsis thaliana						
REFERENCE	1 (bases 1 to 84196)	1 (bases 1 to 84196)						
AUTHORS	Bloeker, H., Mewes, H.W., Lemcke, K., Meyer, K.F.X., Quetler, F. and Salanoubat, M.	Bloeker, H., Mewes, H.W., Lemcke, K., Meyer, K.F.X., Quetler, F. and Salanoubat, M.						

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 84196)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail: lemc@emips.biochem.mpg.de, mayer@emips.biochem.mpg.de, Project Coordinator: Marcel Salanoubat and Francis Queller, Grouperment d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France; http://www.genoscope.cns.fr

On Jan 27, 2000 this sequence version replaced gi:6434247.
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
Location/Qualifiers
1. 84196
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
2983. .3338
/note="355 bp LTR"
3339. .7564
/note="LTR-transposon"
7565. .7922
/note="367bp LTR"
12339. .13197
/gene="T3A5.10"
join(12339. .12665,12742. .12803,12888. .13055,13125. .13197)
/gene="T3A5.10"
/note="similarity to cyclin-dependent kinase inhibitor (ICK1), Arabidopsis thaliana, EMBL:AT094772.1"
/codon_start=1
/product="putative protein"
/protein_id="CAB62432.1"
/db_xref="GI:6561966"
/translation="MAVRRRRRVDVENEYTTTTRKRRKEEVVDLYESRIILSPCY QATNRGIVANSAASGSETVYVRRKDSPPVEQCIIEEDSSVSCSTSESKRR IEVPLEENNDDRETETSWITDINKSESMNDSSVAVEDESRRRLKSLHETV KEALDEDFQVAEKDLNKKLECSMKYNFPEKDEPLGGRYEMVKLP"

12339. .12665
/gene="T3A5.10"
/number=1
12666. .12741
/gene="T3A5.10"
/number=1
12742. .12803
/gene="T3A5.10"
/number=2
12804. .12887
/gene="T3A5.10"
/number=2
12888. .13055
/gene="T3A5.10"
/number=3
13056. .13124
/gene="T3A5.10"
/number=3
13125. .13197
/gene="T3A5.10"
/number=4
13198. .16470
/gene="T3A5.20"
complement(join(15858. .15901,16014. .16470))
/note="similarity to predicted proteins, Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB62433.1"
/db_xref="GI:6561967"
/translation="MARLEMHNSMNGAARISFSENFVEIRSEKSNNAKNNINSSSPS MPADFAFSVTIDYSMPADELFLKGLLPKREKSHVARTIGDELLTBEBCSMYDGNIF

SLRPIILSSSFSFKGTWRELGLKRTHVRSKTKDKVNEVLSODHKIISGNVATREC QVADTR"
complement(15858. .15901)
/number=1
complement(15902. .16013)
/number=1
complement(16014. .16470)
/number=2
18303. .19931
/gene="T3A5.30"
complement(18303. .19931)
/codon_start=1
/product="scarecrow-like 7 (SCL7)"
/protein_id="CAB62434.1"
/db_xref="GI:6561968"
/translation="MAVYCTDSGNLMAIAOOLIKOKOQOOSOHOOQEOEOEPNWPV PSFGTLLSGSGSDPFOVYNTNGFHPLELHONAAVASEEFDSEMEESLINDGAS QPNDRPIYGHDPVSPSPRSASYSILNRVKKDSASQQLPPPPASTAIRGSPSPPO HPPPPPOPDLDNPIFKAIHDYARKETPDLIRIKESVSESGDYIQRGYFAE ALSHKEESPSSSSSSLEDFILSYKTLMDACPYSKEFHLNANQAIIEATNQSNHII VDFGIFQIQWSALQALATRRSGKPTIRISGIPAPSGDSPGLATGRLDFA AILDNFEFYPVLPIQLNGSSFRVDEVLVYNNMELYKLIDETATGTALRLA RSLNRPVYTLGEYVSLNRVEFANRVKSLRFAVAFESLEPNDROSKERLVRVLL FGRIMDLVRSDDDNNKRGTRFGLMEKEQORVLMKAGFEPPVAFSNVAQAKLLIM NNTSTLSVSESEPGFISLANNVPLTVSSMR"
complement(18303. .19931)
/number=1
26093. .28999
/gene="T3A5.40"
complement(join(26093. .26338,27091. .27300,27308. .27386, 27478. .27570,27723. .28037,28136. .28288,28374. .28698, 28779. .28999))
/note="Contains Cytochrome P450 cysteine heme-ligand signature AA455-464: Prokaryotic membrane lipoprotein lipid attachment site AA320-330"
/codon_start=1
/product="steroid 22-alpha-hydroxylase (DWF4)"
/protein_id="CAB62435.1"
/db_xref="GI:6561969"
/translation="MFEETHHTLPLLLPSELSTLLFLILKRRNRKTRNLPFGKS GMPFLGEGTIGYIKPTATTLGDFMOHVSCKGKLYRSNLFGPEPTVSDAGINRILQ NEGRLPFESYPRSTIGGILKSMVLVYGDMDRDRSTISLNFSAARLPTILLKVERH TLPIYDSMOQNSIFSAODEAKKFTFNMAKIKINGMDPEEETEDQLKKRYVFMKGVAS APLNPGTATYKALQSRATILKTERKWEKKLIDKEDEEVEKTEDEAKMSDQ VRKQRTDDDLGWLKHSNLSSTEDILDLISLPSAGHETSSVAIALAIFFLQACPKAV BELREHELEIARAKKEGSELDNDYKMDFTCCVINEITRLGIVVFFLRKAKADV RYKGYDIPSGWKVLPVISAIVHLDNSRYDOPULFNPMDQOONNGSSGSGSFTWGN NYMPFGGGRPCAGSELAKLMAVFIHHLVLKFMWELAEADKPPAFPPVDPNCLPIR VSRLL"
complement(26093. .26338)
/number=1
complement(26339. .27090)
/number=1
complement(27091. .27200)
/number=2
complement(27201. .27307)
/number=2
complement(27308. .27386)
/number=3
complement(27387. .27477)
/number=3
complement(27478. .27570)
/number=4
complement(27571. .27722)
/number=4
complement(27723. .28037)
/number=5
complement(28038. .28135)
/number=5
complement(28136. .28288)
/number=6
complement(28289. .28373)
/number=6

```

exon      complement(28374. .28698)
          /number=7
intron    complement(28699. .28778)
          /number=7
exon      complement(28779. .28999)
          /number=8
gene       38307. .41323
          /gene="r3A5.50"
          complement(join(38307. .38796,38877. .39090,39189. .39270,
40180. .40242,40337. .40401,40489. .40564,40655. .40734,
40805. .40919,41225. .41323))
          /note="Contains Eukaryotic putative RNA-binding region
RNP-1 signature AA179-186"
          /codon_start=1
          /product="U1 snRNP 70K protein"
          /protein_id="CA62436.1"
          /db_xref="GI:6561970"
          /translation="MODSGDPLRNPNAVOARAVONRANVQLKMGSHPTGLTN
NLKLEPPPLEYKPPPEKRCPCPYTGMAQVSNFAPGDEYAPPEVELEPSOKR
ERHHLRLKEGYEKAAEDLKDPNNDPATGDPYKTLFVSLNATESSESKIKREFS
VGRIRVHYLYTQLTNPKRGVAFIEYMTROMKAAYKQADGQKIDGRVLYDVERGRT
VPMWRPRRLGGGLGTSRVGGGEIIVGEOPOGRTSOSSESPRPREERKSKERE
KSRKLSHQPRERSRDRREDKHNRDQGGRRDRDRDRDRDRDRDRDRDRDRDRDR
GNDRTSRDHRDRSRKREDYEGGEYEHGGGRRSRERAEYRGPEERKGYIEDQGG
TDRYSHRYDKMEEDPFYREYERKRSKRSSESEHYA"
exon      complement(38307. .38796)
          /number=1
intron    complement(38797. .38876)
          /number=1
exon      complement(38877. .39090)
          /number=2
intron    complement(39091. .39188)
          /number=2
exon      complement(39189. .39270)
          /number=3
intron    complement(39271. .40179)
          /number=3
exon      complement(40180. .40242)
          /number=4
intron    complement(40243. .40336)
          /number=4

Query Match      97.9%; Score 3134.2; DB 8; Length 84196;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 3186; Conservative 0; Mismatches 13; Indels 4; Gaps 4;

QY 1 ATGTGGTATTATATGTTGGGTCGGTTGAGCTACATATTAATTCGTTTCGGT 60
DB 32199 ATGTGGGATATATATGTTGGGTCGGTTGAGCTACATATTAATTCGTTTCGGT 60
QY 61 TATTCGTTCACATGATTTGAGTTGGTCTCAATTTGATCCAGATTAATTAATTT 120
DB 32139 TATTCGTTCACATGATTTGAGTTGGTCTCAATTTGATCCAGATTAATTAATTT 120
QY 121 AAAATTCATTTAAATTTTACAAGTAATTAATTAATTCCTTACATGATTTGATTAACA 180
DB 32079 AAAATTCATTTAAATTTTACAAGTAATTAATTAATTCCTTACATGATTTGATTAACA 180
QY 181 AATATCATCTTGGTATAGAGAAATATGAGTTGGAATTTATATATTAATTAAGGAA 240
DB 32019 AATATCATCTTGGTATAGAGAAATATGAGTTGGAATTTATATATTAATTAAGGAA 240
QY 241 TAATCATTCATTTGGTGGATTTACACAGTTAATTTTGGTTTCTTTGTTATATGT 300
DB 31959 TAATCATTCATTTGGTGGATTTACACAGTTAATTTTGGTTTCTTTGTTATATGT 300
QY 301 AATTGAGTAAATCAAGAGATTTGATGGAAGTGAATTAACATATTCGTTAGACCCCA 360
DB 31899 AATTGAGTAAATCAAGAGATTTGATGGAAGTGAATTAACATATTCGTTAGACCCCA 360
QY 361 AAAAAAAAAAAAAAAAACAACAAACAAACCCGCCGATATAGTTTTGGTTGGATT 420
DB 31839 CAAAAAAAAAAAAAAAACAACAAACAAACCCGCCGATATAGTTTTGGTTGGATT 420

```

```

QY 421 AGGTTATTTGATCATATTTACATGACATCTTCTTTGATTAATGAAGATTTCTTAC 480
DB 31780 AGGTTATTTGATCATATTTACATGACATCTTCTTTGATTAATGAAGATTTCTTAC 480
QY 481 CAATTAATTTTCGATTCATATCTTGATTTATTAATTAATTCGATTCGATTCAC 31721
DB 31720 CAATTAATTTTCGATTCATATCTTGATTTATTAATTAATTCGATTCGATTCAC 31721
QY 541 GTTTTCGATTCATCTCAGATTCATGATTTATGTTCTGCTTAATCCAGCAATTTTACA 600
DB 31660 GTTTTCGATTCATCTCAGATTCATGATTTATGTTCTGCTTAATCCAGCAATTTTACA 600
QY 601 AGAGTATTTGAGAAAAACCGAAAAATAGAAAAAGGAAAGAGTATGACCCATGAGATG 660
DB 31600 AGAGTATTTGAGAAAAACCGAAAAATAGAAAAAGGAAAGAGTATGACCCATGAGATG 660
QY 661 TGAATTAATTTACAAAGAGATTAAGAGATGACACCAAAAGGTTGTGGAATTAATGTCCT 720
DB 31540 TGAATTAATTTACAAAGAGATTAAGAGATGACACCAAAAGGTTGTGGAATTAATGTCCT 720
QY 721 GCCAGCTTTCTCTGACATCAATATGACACCCATTTGGATTTTCTGATATTCGTTAAAA 780
DB 31480 GCCAGCTTTCTCTGACATCAATATGACACCCATTTGGATTTTCTGATATTCGTTAAAA 780
QY 781 TTTGCGATTAACGATTTGAAAAATATTTTATTTGATGATGATCAATATTTATGTTCCA 840
DB 31420 TTTGCGATTAACGATTTGAAAAATATTTTATTTGATGATGATCAATATTTATGTTCCA 840
QY 841 GGTATTTGCAATATCTTCTGTTTAAAGCATATTTGCTTTCTTTGTTTCTGTTCTCT 900
DB 31360 GGTATTTGCAATATCTTCTGTTTAAAGCATATTTGCTTTCTTTGTTTCTGTTCTCT 900
QY 901 TAACTATATTTATCGGGATATATGATTAACATGATATATCAAAACAAATTTCTGGG 960
DB 31300 TAACTATATTTATCGGGATATATGATTAACATGATATATCAAAACAAATTTCTGGG 960
QY 961 ACCATTTGCAATATCTTCTGTTTCAAAACATTTAGGGACCTGACCTGCACTTAAATTA 1020
DB 31240 ACCATTTGCAATATCTTCTGTTTCAAAACATTTAGGGACCTGACCTGCACTTAAATTA 1020
QY 1021 CGATTTTACAGCGCTGACATGATGATGATTAAGCATTAAGCAAGGACCGCTCAAGC 1080
DB 31181 CGATTTTACAGCGCTGACATGATGATGATTAAGCATTAAGCAAGGACCGCTCAAGC 1080
QY 1081 TATTTATCAAAAGTTACAAAGTATATGATTTAGCTTTGAAATCTTTAGAAAAATTTGGAATTA 1140
DB 31121 TATTTATCAAAAGTTACAAAGTATATGATTTAGCTTTGAAATCTTTAGAAAAATTTGGAATTA 1140
QY 1141 CCGGTTGTTATGTAATATAGATTTAGTGTAAACAATATGTTAATTAAGTGTGCA 1200
DB 31061 CCGGTTGTTATGTAATATAGATTTAGTGTAAACAATATGTTAATTAAGTGTGCA 1200
QY 1201 ACATATACATATTCCTTACAGAAAAACAACTTAAGAGAAGTTAACATATTCATATAT 1260
DB 31001 ACATATACATATTCCTTACAGAAAAACAACTTAAGAGAAGTTAACATATTCATATAT 1260
QY 1261 GGGTATGCTATACCTTTCACGTATGCTATGCTATGAGCTAAGAAATGTTATGATGTC 1320
DB 30941 GGGTATGCTATACCTTTCACGTATGCTATGCTATGAGCTAAGAAATGTTATGATGTC 1320
QY 1321 GATAATTAATTTACACGCGGTGTAATTAATTTAGGACCGTATGTTAGCATCTGCA 1380
DB 30881 GATAATTAATTTACACGCGGTGTAATTAATTTAGGACCGTATGTTAGCATCTGCA 1380
QY 1381 ATATCATCTTCTGTTGTCACAACTAATAAACAACAAAAAGAAACGATTTT 1440
DB 30821 ATATCATCTTCTGTTGTCACAACTAATAAACAACAAAAAGAAACGATTTT 1440
QY 1441 CTGGAATTCATTAATGATCTAATTAATGATGATCTTTTGGGTTACAGTTTGAAGTCC 1500
DB 30761 CTGGAATTCATTAATGATCTAATTAATGATGATCTTTTGGGTTACAGTTTGAAGTCC 1500

```

QY 1501 TCTACACCGGTGACCATCTGCACATTAATTCCTTCTTTAATGACATCTTAACAT 1560
 Db 30701 TCTACACCGGTGACCATCTGCACATTAATTCCTTCTTTAATGACATCTTAACAT 30642
 QY 1561 ATTATTTGTAGTGAATTAATTAAGAGCACTGTACATTAATTAATTAATTAAG 1620
 Db 30641 ATTATTTGTAGTGAATTAATTAAGAGCACTGTACATTAATTAATTAATTAAG 30582
 QY 1621 ATACTAGTATGTATTAATTCACAAATACATCTTGGATGTTAACTTAATCTTCTTCT 1680
 Db 30581 ATACTAGTATGTATTAATTCACAAATACATCTTGGATGTTAACTTAATCTTCTTCT 30522
 QY 1681 TCCACCGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
 Db 30521 TCCACCGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 30463
 QY 1741 ATGAAGATTAACCTTAATGATCTTAATTTTGAATTAATTAATTAATTAATTAATTAAT 1800
 Db 30462 ATGAAGATTAACCTTAATGATCTTAATTTTGAATTAATTAATTAATTAATTAATTAAT 30403
 QY 1801 AATTAACCGTATGTTTGTGTCATTAATGACAGCTCTACACCTGTGATGATCAATTTT 1860
 Db 30402 AATTAACCGTATGTTTGTGTCATTAATGACAGCTCTACACCTGTGATGATCAATTTT 30343
 QY 1861 TCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1920
 Db 30342 TCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 30283
 QY 1921 ACATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980
 Db 30282 ACATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 30223
 QY 1981 AATATGATGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040
 Db 30222 AATATGATGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 30163
 QY 2041 TAAATACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100
 Db 30162 TAAATACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 30103
 QY 2101 AATATCTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160
 Db 30102 AATATCTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 30043
 QY 2161 TTGTTACAGAACTTCCAAATTTTCTTTTATGACAGAAATTAACAGATGAGAAA 2220
 Db 30042 TTGTTACAGAACTTCCAAATTTTCTTTTATGACAGAAATTAACAGATGAGAAA 2280
 QY 2221 CTATTTTGTGGAATGGAAGTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280
 Db 29982 CTATTTTGTGGAATGGAAGTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280
 QY 2281 AGCCTATACGCGCTCAAGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2340
 Db 29922 AGCCTATACGCGCTCAAGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2340
 QY 2341 TCAGATTTGGAGCAACAATGAAGAAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2400
 Db 29862 TCAGATTTGGAGCAACAATGAAGAAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2400
 QY 2401 TTGAGTAAATGTTTCTGCTATGAGGGGCAAAAAAGACAAATCCCAAGTCTAC 2460
 Db 29802 TTGAGTAAATGTTTCTGCTATGAGGGGCAAAAAAGACAAATCCCAAGTCTAC 2460
 QY 2461 GGGTTTGAATCTGCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
 Db 29742 GGGTTTGAATCTGCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
 QY 2521 TAGGGGTCTTCTACATTTTCTACGTTCTACCCCTACTGCTGAGCCGACCCCTTTTCCCA 2580
 Db 29682 TAGGGGTCTTCTACATTTTCTACGTTCTACCCCTACTGCTGAGCCGACCCCTTTTCCCA 2580
 QY 2581 TATCTTAAGGTATTTTGGAAATCCCAATTTAAACCAATGAGACCGTACCGGACTTCC 2640

Db 29622 TATCTTAAGGGTAAATTTTGGAAATCCCAATTTAAACCGATTGAGACCGTACCGGACTTCC 29563
 QY 2641 TGGGATTTCTGCTGAGACATTTATCAAAAAATTTATGACAGATGGGTTTATTAATTTAA 2700
 Db 29562 TGGGATTTCTGCTGAGACATTTATCAAAAAATTTATGACAGATGGGTTTATTAATTTAA 29503
 QY 2701 AACTCACAACCTTGATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
 Db 29502 AACTCACAACCTTGATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
 QY 2761 ATCTATGACCTTTTCTTTTCTACACCGGTGATGAAGTTATAGTATTAATTAATTAATTAATTAAT 2820
 Db 29442 ATCTATGACCTTTTCTTTTCTACACCGGTGATGAAGTTATAGTATTAATTAATTAATTAATTAAT 2820
 QY 2821 ACAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
 Db 29382 ACAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
 QY 2881 CTATTTGACGATCGAGCTTCTGCAACTTTTGTTTTAAATTAATTAATTAATTAATTAATTAATTAAT 2940
 Db 29322 CTATTTGACGATCGAGCTTCTGCAACTTTTGTTTTAAATTAATTAATTAATTAATTAATTAATTAAT 2940
 QY 2941 AAGTATTAAGAGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2993
 Db 29262 AAGTATTAAGAGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2993
 QY 3001 TGAACCTGATGTTTGA-TTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3059
 Db 29202 TGAACCTGATGTTTGA-TTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3059
 QY 3060 GATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3119
 Db 29142 GATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3119
 QY 3120 TTTGATTAAGTCCGATCCCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3179
 Db 29082 TTTGATTAAGTCCGATCCCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3179
 QY 3180 AGAGAGAGAGAACTAGCTCC 3202
 Db 29022 AGAGAGAGAGAACTAGCTCC 3202
 RESULT 2
 AF044216 4818 bp DNA linear PLN 25-JUN-2001
 LOCUS
 DEFINITION Arabidopsis thaliana steroid 22-alpha-hydroxylase (DMF4) gene,
 complete cds.
 ACCESSION AF044216
 VERSION AF044216.1 GI:2935341
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana.
 Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 4818)
 Choe,S., Dikes,B.P., Fujioke,S., Takatsuto,S., Sakurai,A. and
 Feldmann,K.A.
 The DMF4 gene of Arabidopsis encodes a cytochrome P450 that
 mediates multiple 22alpha-hydroxylation steps in brassinosteroid
 biosynthesis
 JOURNAL Plant Cell 10 (2), 231-243 (1998)
 MEDLINE 98158690
 PUBMED 9490746
 REFERENCE
 2 (bases 1 to 4818)
 Choe,S., Dikes,B.P., Azpiroz,R. and Feldmann,K.A.
 Direct Submission
 Submitted (22-JUN-1998) Plant sciences, University of Arizona,
 Tucson, AZ 85721, USA
 JOURNAL
 FEATURES
 source location/Qualifiers
 1..4818

gene
/organism="Arabidopsis thaliana"
/strain="Ws-2"
/db_xref="taxon:3702"
/chromosome="III"
/map="8 cm from nga6"
990..4398
/gene="DMF4"
/note="DMF4"
/allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4"
990..1055

TATTA_signal
/gene="DMF4"
join(<1133..1353,1434..1758,1844..1996,2095..2409,
2562..2654,2746..2824,2931..3040,3795..4398)
/gene="DMF4"
/product="steroid 22-alpha-hydroxylase"
join(1133..1353,1434..1758,1844..1996,2095..2409,
2562..2654,2746..2824,2931..3040,3795..4040)
/function="Prassinosteroid biosynthetic enzyme"
/note="member of the cytochrome P450 superfamily; CYP90B1"
/product="steroid 22-alpha-hydroxylase"
/protein_id="AAC05093.1"
/db_xref="GI:2935342"

CDS
/translation="MFEHHHTLPLILPLSLILLILKRNKRTRNLPKGS
GMPFLEGTIGLKPYTATTLGDMOQHSKTKGKTKYRSITFEPTVSDAGLNPLIO
NEGRLEFESQYRSTIGILKSMVLVGDMDKMSISLNLSHARKETILKDEVERH
TLFVLDISWQNSRISADDEAKTFENLAKHINMSDPEETBOLKEKYEYFMGYS
APLNPGRATKHALQSRTILKTERKEERKDIKEDEQEEVETEDEREMSSH
VRKRTDDELKAWLKNSLSTEQILDLILSLPAGHETSSVALAIFLPOACPKRV
BELREHLEIARAKKEGESELDNDYKMDFTQCINETLGNVRFILKRLKY
RYKGYDIPSGKVLPIVSAVHLNDSRYDQPLFPMWMOONNASSSGSSSTWEN
NMPFGGPRLCAGSELAKEMAVFIHIVLKEFWELAEEDDPFAFFVDPNGLPLR
VSRIL"

BASE COUNT 1566 a 727 c 888 g 1637 t
ORIGIN

Query Match 35.4%; Score 1132; DB 8; Length 4818;
Best Local Similarity 100.0%; Pred. No. 1.1e-153;
Matches 1132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2071 GAATCAATATTCCTTAAATATTTTGAATCTACAAATTTATATTTAGTCAAT 2130
DB 1 GAATCAATATTCCTTAAATATTTTGAATCTACAAATTTATATTTAGTCAAT 60
QY 2131 AACATGCAATGAAAGTCCAAAAAATTTGTTACAGAAAGTCCAAATTTT 2190
DB 61 AACATGCAATGAAAGTCCAAAAAATTTGTTACAGAAAGTCCAAATTTT 120
QY 2191 TTATGGAACAGAAATACAGATAGAAACTATTTGTTGGAATGGAATAGTATA 2250
DB 121 TTATGGAACAGAAATACAGATAGAAACTATTTGTTGGAATGGAATAGTATA 180
QY 2251 TACATTAAGCAAAATTTAAAAAATTTATATTAAGCTATACGCGCTCAAGATGTATCTA 2310
DB 181 TACATTAAGCAAAATTTAAAAAATTTATATTAAGCTATACGCGCTCAAGATGTATCTA 240
QY 2311 GTAGGTATTTATATATGATGCGATTCAGAAATGGGACACAAATGAAAGCAAT 2370
DB 241 GTAGGTATTTATATATGATGCGATTCAGAAATGGGACACAAATGAAAGCAAT 300
QY 2371 TAAAAATTAATTTAAAAAATTTATATTAAGTAAAGTCTTTTTCGACTATTGAGG 2430
DB 301 TAAAAATTAATTTAAAAAATTTATATTAAGTAAAGTCTTTTTCGACTATTGAGG 360
QY 2431 GGCAGAAAAAGCAATGCCAAAGTCTACGGGTTTGACGTGCCAATGGTAAATATCT 2490
DB 361 GGCAGAAAAAGCAATGCCAAAGTCTACGGGTTTGACGTGCCAATGGTAAATATCT 420
QY 2491 AATACTCTGTCTTTGACGCAACGCGGTGAGGGTCTCTTGACATTTTACGTCTCT 2550
DB 421 AATACTCTGTCTTTGACGCAACGCGGTGAGGGTCTCTTGACATTTTACGTCTCT 480

QY 2551 ACCCTACTGCTGAGGCCACCTTTTCCATATCTTAAGGGTAAATTTGAAATGCCAT 2610
DB 481 ACCCTACTGCTGAGGCCACCTTTTCCATATCTTAAGGGTAAATTTGAAATGCCAT 540
QY 2611 TTAACCGATGAGACCCGACCGACTCTCTGGATTTCTGTGAGCATTTATCAAAAT 2670
DB 541 TTAACCGATGAGACCCGACCGACTCTCTGGATTTCTGTGAGCATTTATCAAAAT 600
QY 2671 TATTAGCAGAAATGGGTTTATTTAAAAACACAACTGATCAATATTTAT 2730
DB 601 TATTAGCAGAAATGGGTTTATTTAAAAACACAACTGATCAATATTTAT 660
QY 2731 AAACCTTTTACGATGATGCTGACGATCTATCAATGACTTTTCTTACACAGCT 2790
DB 661 AAACCTTTTACGATGATGCTGACGATCTATCAATGACTTTTCTTACACAGCT 720
QY 2791 GGATGAAAGTTATAGTACTATTAGCAGACAAATTTGATATATATCATTAATCC 2850
DB 721 GGATGAAAGTTATAGTACTATTAGCAGACAAATTTGATATATATCATTAATCC 780
QY 2851 ATGATTTTATGATATTAATATAGCTTTAACTATTTCAGATCCAGCTTTCTGCACT 2910
DB 781 ATGATTTTATGATATTAATATAGCTTTAACTATTTCAGATCCAGCTTTCTGCACT 840
QY 2911 TTGTTTTTAAATTTAAAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2970
DB 841 TTGTTTTTAAATTTAAAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
QY 2971 AAAGTAAATGAAACAGGAGAAACAAAGCCATGAGTCTATGTTTATTAAGCTTAATA 3030
DB 901 AAAGTAAATGAAACAGGAGAAACAAAGCCATGAGTCTATGTTTATTAAGCTTAATA 960
QY 3031 AGAATTTTATTTAAATTTAAATGACATGATTAACAATTTATTTCTGACTTTTAA 3090
DB 961 AGAATTTTATTTAAATTTAAATGACATGATTAACAATTTATTTCTGACTTTTAA 1020
QY 3091 ACCCTCTTACAAACAGAGTCCCTTTTCAATGATGATGATGATGATGATGATGATGAT 3150
DB 1021 ACCCTCTTACAAACAGAGTCCCTTTTCAATGATGATGATGATGATGATGATGATGAT 3210
QY 3151 ACAAGCCATTACAAAGAGAAATGAGTGAAGAGAGAGAGAAATAGTCTCC 3202
DB 1081 ACAAGCCATTACAAAGAGAAATGAGTGAAGAGAGAGAGAAATAGTCTCC 1132

RESULT 3
AL844732
LOCUS
DEFINITION
Arabidopsis thaliana transposon insertion STS SM.3.29533, sequence tagged site.
AL844732
ACCESSION
AL844732.1 GI:21998461
VERSION
STS, STS, sequence tagged site.
KEYWORDS
Arabidopsis thaliana.
SOURCE
Arabidopsis thaliana
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
Clarke,J.H., Bowles,B., Carter,J., Hart,D., McCullagh,B., Murphy,G., Langham,S., Leclerc,C., Jones,J.D.G. and Bevan,M. Unpublished
2 (bases 1 to 191)
REFERENCE
Clarke,J.H.
TITLE
Direct Submision
JOURNAL
Submitted (25-JUL-2002) Clarke J.H., John Innes Centre, Colney Lane, Norwich, NR4 7UJ, UK
COMMENT
AT denotes an activation tag dissociation transposon within a single line, ET an enhancer trap dissociation transposon, GT a gene trap dissociation transposon, MT a mis-expression enhancer trap dissociation transposon, SM a defective suppressor mutator transposon, _3 denotes a sequence derived from the 3' end of the transposon, _5 denotes a sequence derived from the 5' end of the

QY 1065 AAGGACCGCTCAGCTATTTATACAAAGTACAACTGATATAGCTGGAATCCTTAA 1124
 DB 73404 AAT 73463
 QY 1125 GAAATTTTGGATTTACCGGCTGTTATGTAATATATAGATTTAGTGTAAACAAATATGTT 1184
 DB 73464 ATTAAT 73523
 QY 1185 AATCAATAGTGTCAACATATACATATATCTTACGAAATTAACAACTTAAGACAGT 1244
 DB 73524 AATTAATTAACAAAT 73583
 QY 1245 TAACATATGCAATATATGGAATGCTATACCT-----TTCAGCTATGCTATAGTA 1295
 DB 73584 TAT 73643
 QY 1296 GACTAAAGAT 1355
 DB 73644 AATTAATTAACAAAT 73703
 QY 1356 GGACCGTATGCTACATGCAATATATATATATATATATATATATATATATATATATAT 1415
 DB 73704 TAT 73762
 QY 1416 ACAAGAAAAAAGAAACGATTTTCTGATTCATTCATATGATCTAAATGCAATAGAT 1475
 DB 73763 TAT 73822
 QY 1476 CTTTGGGTATACAGTTTGCAGATCCTTACAAAGCTGTAAACATGCAATATATATAT 1535
 DB 73823 AATATGTTAT 73882
 QY 1536 GCTTTCTTAAATGCAATCTTAAATATATATATATATATATATATATATATATATATAT 1595
 DB 73883 TAT 73942
 QY 1596 TCTAAT 73942
 DB 73943 TATTAAT 74002
 QY 1646 ACATATCTTGTATGTTTAAATCTTATCTTCTCTCTACGGATATATATATATATATAT 74002
 DB 74003 AAT 74062
 QY 1706 GAGTAAATAA--AAGTTTGTCTTATATATATATATATATATATATATATATATATAT 1763
 DB 74063 AATTAAT 74122
 QY 1764 TAAATTTTGAATAATGTAACCTTTTACATGATTAATTAACCGTATGTTTGTGTC 1823
 DB 74123 TAAATTTTGAATAATGTAACCTTTTACATGATTAATTAACCGTATGTTTGTGTC 1823
 QY 1824 ATAAATGACAGCTCTACAGCTGATGATCAATTT--TTTCGCAATATATATATATATAT 74182
 DB 74183 AATTAAT 1880
 QY 1881 AATCAATGCTACATCAAT 74242
 DB 74243 ATTAAT 1940
 QY 1941 AATTTTGAATAATGTAACCTTTTACATGATTAATTAACCGTATGTTTGTGTC 74302
 DB 74303 TAAATTTTGAATAATGTAACCTTTTACATGATTAATTAACCGTATGTTTGTGTC 74362
 QY 2001 TTTCCATATGTTCTTAAAT 2060
 DB 74363 TTAATTAAT 74422
 QY 2061 AATAGTGTGCAATCAAT 2116
 DB 74423 TAAATTTGAATAATGTAACCTTTTACATGATTAATTAACCGTATGTTTGTGTC 74482

QY 2117 AATATATGAT 2176
 DB 74483 TATTTTAAAT 74542
 QY 2177 CCAATTTTATTTTAT 2236
 DB 74543 ATTAAT 74602
 QY 2237 TCGAAT 2296
 DB 74603 TAT 74662
 QY 2297 AAGTAT 2356
 DB 74663 ATTAAT 74720
 QY 2357 AATGAAACGGAATTAAT 2404
 DB 74721 ATTTTAAATTAAT 74768

RESULT 5
 AC004157
 LOCUS
 DEFINITION
 AC004157 169546 bp DNA linear HTG 12-AUG-2000
 PROGRESS ***, 2 unordered pieces.
 AC004157
 VERSION
 AC004157.8 GI:9797712
 KEYWORDS
 HTG; HTGS_Phasel.
 SOURCE
 ORGANISM
 Plasmodium falciparum.
 Plasmodium falciparum.
 Eukaryota; Alveolata; Apicomplexa; Hemosporida; Plasmodium.
 1 (bases 1 to 169546)
 Hymen, R.W., Fung, E.L., Qin, F., Rowley, D., Mao, J., Tamaki, T.,
 Kuri, O.B., Conway, A.B. and Davis, R.W.
 Plasmodium falciparum 3D7 chromosome 12
 Unpublished
 2 (bases 1 to 169546)
 Hymen, R.W., Fung, E.L., Qin, F., Rowley, D., Mao, J., Tamaki, T.,
 Kuri, O.B., Conway, A.B. and Davis, R.W.
 Direct Substitution
 Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology
 Center, Stanford University, 855 California Avenue, Palo Alto, CA
 94304, USA

COMMENT
 On Aug 12, 2000 this sequence version replaced gi:8810447.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source
 1. 169546
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /chromosome="12"
 /clone="PRYAC293"
 /clone="3D7"
 Location/Qualifiers
 * 1 23466: contig of 23466 bp in length
 * 23467 23666: gap of unknown length
 * 23667 169546: contig of 145880 bp in length.

BASE COUNT 69871 a 15381 c 15705 g 68389 t 200 others
 ORIGIN

Query Match 4.2%; Score 133.2; DB 2; Length 169546;
 Best Local Similarity 44.7%; Pred. No. 1.5e-10;
 Matches 1041; Conservative 0; Mismatches 1248; Indels 39; Gaps 12;
 QY 108 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 166
 DB 84033 ATTAAGCTATTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 84092

	Db	84663	ATTATATATCTTTTAAATTAATAAATAATTAATGTTTAAATATATATTAATTAATTA	84604
QY	1829	GACAGCCCTCACAAGTGATGCATGTCATTTTTCTCGAAATATTAATTTGAATTCAT	1888	
Db	84603	TTTAAATATCTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	84549	
QY	1889	GCATCATCAATAGAGAACAACAGCTGATTTACATTTTAAATTAAGAACAATTTTG	1948	
Db	84548	ATTATATATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	84489	
QY	1949	AAAAATGTTTAAATTTCTAACAAATATATTAATAATATGATGCCATATATGTTCTAT	2008	
Db	84488	ATTATTTTATTTTATTTTATTAATAATTAATTAATTAATTAATTAATTAATTAAT	84429	
QY	2009	GTTCTTAAAATTAATTTTTTTTTTAAATTTAGTTATTAATATACATTAATGAACCAATATAGTTG	2068	
Db	84428	TTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	84370	
QY	2069	GTCATTCACAAATTCGCCATTAATTTTGTGAATCTCACAAATTAATTAATTTAGTCA	2128	
Db	84369	TAAATTAATTAATTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	84310	
QY	2129	ATAACAATGCATAGAAAAGTCCAAAAAAATTTGGTTAACAGAACTCCAAATTTTTTT	2188	
Db	84309	ATTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	84250	
QY	2189	TTTTTTGGAACAAGAAATTAACAGATAGAAAACATTTGTTGTCGAATGAGTAGTAA	2248	
Db	84249	ATTTAATTTATTTATTTATTTTATTTTATTAATTAATTAATTAATTAATTAATTAAT	84195	
QY	2249	TATACATTAAGCAAATTTTAAAAAATTATATA	2280	
Db	84194	TATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	84163	
RESULT 8				
AEO01398				
LOCUS				
DEFINITION		AEO01398	14867 bp DNA linear INV 16-APR-2002	
ACCESSION			Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.	
VERSION		AEO01398	AEO01362	
KEYWORDS		AEO01398.1	GI:3845197	
SOURCE				
ORGANISM			Plasmodium falciparum 3D7.	
REFERENCE			Plasmodium falciparum 3D7.	
AUTHORS			Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
			1 (bases 1 to 14867)	
			Gardner,M.J., Retelien,H., Garuci,D.J., Cummings,L.M., Arayind,L., Koonin,E.V., Shalimov,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,R., Jing,X., Aston,C., Lai,Z., Schwartz,D.C., Bertea,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.	
TITLE			Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum	
JOURNAL MEDLINE			Science 282 (5391), 1126-1132 (1998)	
PUBMED			99021743	
REFERENCE			9804551	
AUTHORS			2 (bases 1 to 14867)	
TITLE			Gardner,M.J.	
JOURNAL			Direct Submission	
FEATURES			Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA	
SOURCE			Location/Qualifiers	
			1..14867	
			/organism="plasmodium falciparum 3D7"	
			/strain="3D7"	
			/db_xref="taxon:36329"	
			/chromosome="2"	
			/complement(1570..2424)	
			/gene="PFB0490C"	
			/complement(1570..2424)	
			/gene="PFB0490C"	
gene				
CDS				

[illegible]

OY	1455	AAGACGCTMAAAGCGATAGACTCTTTGGGTACAAGTTCGAAGTCCCTCAACAGCTGTA	1514
Db	24312	TATAATATATTATTTATATTTTATACAGTTAAATAAATATTTATATTTTATACAGTTATA	24371
OY	1515	ACCATCGCAACTATTAATATGGCTTCTCTTAATGCATCCTTAACATATTTATTTAGTT	1574
Db	24372	TAAATATATTTATATTTTATACAGTTATATTA - ATATATTTATATTTTATACAGTTATAT	24430
OY	1575	GGAATTTAATAGAGCACTGTAAACATTACATATTTATTTAGATACAGTAGTGCA	1634
Db	24431	AAATATATTTATATTTTATACAGTTATATTAATATATTTATATTTATACAGTTATATTA	24490
OY	1635	TATTCOCAATACATCCTTGGATGTTTAACTTAATCTTGTTCTTCCACGGTATAA	1694
Db	24491	ATA-TATTTATATTTATACAGTTATATTAATATTTATAGTTTTATACATTTATATAA	24549
OY	1695	TATTAATCATCGAGTAAAAAAAGTTTGTCTATTTTGGCAGTGAAGAAGATTAACC	1754
Db	24550	TATATTTATATTTTATACATTTGTATTTAATATNTT---ATATATTAATATATTTTA	24606
OY	1755	TATAGCTTTAATTTTTTGAAAATGTAAACCCTTTTACATAGATTAAATACCGTATGT	1814
Db	24607	TATTTATATATTTTATATATTAATATATATATATATATTTATATATATATATATATTA	24666
OY	1815	TTTGTGCCATATGACACACCTCTACACGTGATGCOATTTTTTTTGGCAAAATTTAA	1874
Db	24667	TATATTTGATTTTAAATATCTTTTATATATATATATATATATTTATATTAATATATGTTTA	24726
OY	1875	ATTAGGAATTCAGTCTACTATCATATAGAAAGAACGAGATTTACATTTTAAATTTAA	1934
Db	24727	TATATCACATATATATTTATATTTAATATATATTTTACATTAATAATATATATATTTGA	24786
OY	1935	AGACAAATATTTTGA AAAATGTATAATTTCTAACAAATTTTATAAATATGATGCCAT	1994
Db	24787	TATATATTTATATTTGATATATATATTTATATATATTTAAATATTTTACATTTGATATATAT	24846
OY	1995	AATGATTTCCATGTCTCTTAATAATATTTTTTTTATTTATTTAGTATAATCATATGTA	2054
Db	24847	TTTATATATATTAATATATTTTACATTTGATATATATATTTATATATATTTATTTACA	24906
OY	2055	ACCATA-ATAGTGGTGAATTCCAAATATCTCCATTAATATTTTTTGAAATGTACAAAT	2113
Db	24907	TTTGATTTATTTTATATATATATTTATATATTTATATTTATTTGATATATGTTTATAGATTAAT	24966
OY	2114	ATTATATTTTATGTCATATACATGATGATGAAGTTCCAAAAAAATTTTGTATACAGAAA	2173
Db	24967	ATAATTTTACGTTTAAATATATATTTATATATATATATATTTACGCTTAAATATATATTT	25026
OY	2174	CTTCCAATTTTTTTTTTATATGAAACAAGAAATATACAGATGAAAGACTATTTGTGTG	2233
Db	25027	ATCTATTAATATATATTTATGTTTAAATATATTTATATGATTAATATATATTTACGTTTA	25086
OY	2234	GAATGAGATGATATATACATTTAGCAAAATTTTAAAAAT 2274	
Db	25087	ATATATATTTTACGTTAAATATATATTTTACGTTTATATATAT 25127	

RESULT	10
PFMAL3P5	
LOCUS	86827 bp DNA linear INV 04-MAY-2000
DEFINITION	Plasmodium falciparum MAJ3P5 complete sequence.
ACCESSION	AF034556 AF008971 AF008972 AF008978 AF010141 AF010153 AF010162 AF010206 AF010210 AL139179
VERSION	AL034556.3 GI:7711064
KEYWORDS	HG; centromere; CYP protein; initiation factor E4; Serine/threonine protein phosphatase.
SOURCE	Plasmodium falciparum 3D7. Plasmodium falciparum 3D7
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE	1 (bases 1 to 86827)
AUTHORS	Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T., Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T., Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holroyd,S.,

TITLE
JOURNAL MEDLINE
PUBMED
10448855
2 (bases 1 to 86827)
Bowman, S., Skelton, J., Churcher, C., Lawson, D., Quail, M. and Barrell, B. G.
The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*
Nature 400 (6744), 532-538 (1999)

REFERENCE
JOURNAL
ADTHORS
TITLE
JOURNAL
COMMENT
On or before May 14, 2001 this sequence version replaced
g1:2982535, g1:2982536, g1:2894454, g1:2982554, g1:2982562,
g1:2894469, g1:2982572, g1:2982574, g1:4493931.
For more information about this sequence or the Malaria Project,
see <http://www.sanger.ac.uk/projects/P-falciparum>.

FEATURES
Source
1..86827
/organism="Plasmodium falciparum 3D7"
/strain="3D7"
/db_xref="taxon:36329"
/chromosome="3"
/clone="MAL3P5"
join(322..603,826..1023,1197..1301,1458..2942)
/gene="PFC0575W, MAL3P5.1"
join(322..603,826..1023,1197..1301,1458..2942)
/note="PFC0575W, MAL3P5.1"
/note="PFC0575W (MAL3P5.1), Hypothetical protein, len: 689
aa, possible signal sequence, revised: added new exon 2"
/codon_start=1
/product="Hypothetical protein, PFC0575W"
/protein_id="CAB38969.2"
/db_xref="GI:7711065"
/db_xref="SPRMBL:O97258"
/translation="MYLKNYVYISSCFILDCFSFHLKMKYKNNMNMKSTFFL
RSPOYIKRRKRRIKIVSFRKKQKFLFENLKSGFSGEWRNQYDQYIDVIS
NINLVRIKQVTHKKSEFTKNIKQILHGVSKIDFTINNLVYIKFQMSNT
VHSILNQISSEKVEKDAENYLAHLFLKIDENITLFSMHIMDFEKSQKVICID
IKSKRRKMLSIYINICTLIYFCMLIKYISHLCLFEFFFCFLCYILIR
IYECVGLIRKRIERYNYLCEKKIKFKHMDAKKMEIMKODDLYENHYDELALC
FTMKINTERNNKNTIRSNYDNINNDISIDKDMYNNINIDVNNISIDKIKBEFNP
DDEKLEKLTDEYFOLFNDNITIKYIEEDPLVININDSNINNNININIKKHKIK
DTYNDDDYDEKEEDLVQKNIDYIKNTIGNKKLEEFKNQIFQADIEFNLS
NVNLDQGRVKSNDENKSTENKNTINKGYTELQOMENEFKQADINDNISND
ISNNEINIKIKLNOSDEDINLSDLIERLRKIVLYOKIELYKFOYDIIINQ
YPIIKNEKTVLIDLLNYGKIVMSPDVNSLFEKTKIDISINENKKNQNMKNKKNY
TIMENW"
604..609
/gene="PFC0575W, MAL3P5.1"
/note="predicted splice donor sequence for exon 1 of
PFC0575W (revised)"
816..825
/gene="PFC0575W, MAL3P5.1"
/note="predicted splice acceptor sequence for exon 2
(revised of PFC0575W)"
1024..1029
/gene="PFC0575W, MAL3P5.1"
/note="predicted splice donor sequence for exon 2 of
PFC0575W (revised)"
complement(3354..6644)
/gene="PFC0580C, MAL3P5.2"
complement(3354..6644)
/gene="PFC0580C, MAL3P5.2"
/note="PFC0580C (MAL3P5.2), Hypothetical protein, len:

1097 aa, possible signal sequence, predicted using
hexon"
/codon_start=1
/product="Hypothetical protein, PFC0580C"
/protein_id="CAB38968.1"
/db_xref="GI:4493932"
/db_xref="SPRMBL:O97257"
/translation="MLGILKRNVPYLLVSPSLFAYFLKRRKNDENNVEYILNNNDIE
KIKRIHNNKCSYIPDLFLNIYDSYIKNKLIMLYEFRRKRDKEEYIYTNVVK
KRAIRKYNFSDQENKFEYIYEIVYELVSKIGLISPLIKNSHGHV
SLYYNKLNDKNNLINKEKLYFQINNEHQOAPTNHTHHNNNNNNKPPDIIHSC
KNWISSSYATNNNEKERNINIDYNNINHTNTEKISISIKDNCQHTLIMLYLNOTY
DNICRLNNTNITINFMILNLYCTKMEILLNINHIEMKKTKINOKNNTNS
LEKYSISFEFEKEENHLYDEEQMMNHLKENDKFNYSNENTHNNYKYSIDN
FYDHINSSNRCSPFNKQDOTDNTNHIINGKERYPMKSDHEKNNKGNINIE
DOKDILKIFYLGNKLDIOILNELVYMLHLEFCSKLISIKNHILKCKDE
FDKDKIITYLSADYNNLRNRIILKRESKNERNNINSFASPLSKNITLFEFDEIG
RSRDTYNNYDKEFKTJTNNNNNNDNNDICGNDHICSNNDHICSNNDHICSN
NNNNICSNNNNNICSNNNNNICSNMULDECOBKFNNDYNTKKERKRI
YELAKITYNNIFDYLGKRRKQHEDNTINLYIKKRPWIFLNKIKNDTSFIEH
NNNIVNDIKNNNIIFKKKYLFPSSISYIYIDYVYKLYIYDNLIKKFCR
YELAMNHNINKLTKMKRAFYIYNFDOFYNNYHIIHKKNHKKIHLKCKDE
IDYKFDLYCMNININIFSYIHKYDHCYVRIKRAVKILLYENYKKEENY
YKRIKRYLYNNINISNDLYPNYSIKIYNQKRYKRNKFSIFSLKNDHLLP
LYTORIONCDIESYITKINENKFNPLLYELHYISEKRRKKKFSISSSP
YSYDVMNSFFSYFFSLSYLLFLYHPDMVASYIFFKLTYSGLPTYYSLYNNI
MVCGPKT"
join(7669..7717,7800..7829,7912..7940,8064..8106,
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,
9369..9505,9613..9838)
/gene="PFC0581W"
join(7669..7717,7800..7829,7912..7940,8064..8106,
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,
9369..9505,9613..9838)
/gene="PFC0581W"
/note="PFC0581W, hypothetical protein, len: 324 aa,
revised: new gene prediction, splicing prediction very
tentative"
/codon_start=1
/product="Hypothetical protein, PFC0581W"
/protein_id="CAB90285.1"
/db_xref="GI:7711066"
/translation="MKKKKHYFSIKRVNNDKKNVYVCSKDYIINIMGLIFRL
QKIFHIFANBQIASFSFESHKNTVRVEDIIDEKQWETIDYLSIESIKQNDPS
NDIKTFTYAKNNKQDLITTSYKSKIDYVNASQOMQSPYILNKFSTRMSPPA
LKVDEEIVAKKNKNSFALSNDNSYTKYIVDLTDNIIESETYKNSFASVKKVY
TIKKERKINRRLILSEKYPNMQVMDMEKRIHITFTVINLFFLSISHRHDSV
ONFLKEKNNSDKIQDDIDDEBEKYPBEILREAKKSEEDNDEBL"
7790..7799
/gene="PFC0581W"
/note="potential splice acceptor sequence for exon 2 of
PFC0581W"
7830..7835
/gene="PFC0581W"
/note="potential splice donor sequence for exon 2 of
PFC0581W"
7901..7911
/gene="PFC0581W"
/note="potential splice acceptor sequence for exon 3 of
PFC0581W"
7941..7947
/gene="PFC0581W"
/note="potential splice donor sequence for exon 3 of
PFC0581W"
8055..8063
/gene="PFC0581W"
/note="potential splice acceptor sequence for exon 4 of
PFC0581W"
8107..8112
/gene="PFC0581W"
/note="potential splice donor sequence for exon 4 of
PFC0581W"
8293..8301

```
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 5 of
PFC0581w"
8369..8374
/gene="PFC0581w"
/note="potential splice donor sequence for exon 5 of
PFC0581w"
8559..8567
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 6 of
PFC0581w"
8642..8647
/gene="PFC0581w"
/note="potential splice donor sequence for exon 6 of
PFC0581w"
8704..8712
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 7 of
PFC0581w"
8813..8818
/gene="PFC0581w"
/note="potential splice donor sequence for exon 7 of
PFC0581w"
8913..8923
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 8 of
PFC0581w"
9007..9012
/gene="PFC0581w"
/note="potential splice donor sequence for exon 8 of
PFC0581w"
9112..9121
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 9 of
PFC0581w"
9239..9264
/gene="PFC0581w"
/note="potential splice donor sequence for exon 9 of
PFC0581w"
9361..9368
misc_feature
4.0%; Score 127.8; DB 3; Length 86827;
Best Local Similarity 44.0%; Pred. No. 1.1e-09;
Matches 1051; Conservative 0; Mismatches 1307; Indels 29; Gaps 11;
Query Match
35 TACAATATAATTCGTTCTGTTCTGTTATCTGTTACAGATTTGAGTTGTTCTCAA 94
Db 38839 TAATATACAAATTAATTCATGTTTAAATTAATTAATTTTAAATATATATA 38898
QY 95 TTGGATTCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 154
Db 38899 TATAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 38958
QY 155 ATCTTACATTTGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 214
Db 38959 ATACATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39018
QY 215 TTGGATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 274
Db 39019 AATTTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39078
QY 275 GTTTTGTGTTCTTTGTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 334
Db 39079 ATGTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39138
QY 335 GTAAACATTTGTTATGACCCCAAAAAAATTAATTAATTAATTAATTAATTAAT 394
Db 39139 ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39198
QY 395 CCGATATATGTTTGTGTTCTGTTAGTTTATTTGATTAATTAATTAATTAATTA 454
Db 39199 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 39258
```

```
QY 455 TTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 514
Db 39259 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39315
QY 515 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 573
Db 39316 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39375
QY 574 TTGCTATATCCAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 633
Db 39376 ATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39435
QY 634 GGAAGAGTATGACCCATGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 693
Db 39436 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39495
QY 694 CCAAGAGTGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 753
Db 39496 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39555
QY 754 TTGGATTTCTGATATTCGTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 813
Db 39556 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39611
QY 814 GTTAGCTATCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 873
Db 39612 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39671
QY 874 TTGCTTTCTTTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 933
Db 39672 AATGTTTCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39731
QY 934 TGAATATCAAAACATTTGTTGGACATTTGGAATTAATTAATTAATTAATTAATTAAT 993
Db 39732 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39791
QY 994 GGGACCTGAGCTGACCCCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1053
Db 39792 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39851
QY 1054 CATTAAGCATTAAGACCCGTTCAAGCTTAATTAATTAATTAATTAATTAATTAATTAAT 1113
Db 39852 AATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39911
QY 1114 GAATCTTTAGAAATTTTGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1173
Db 39912 TTAAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39968
QY 1174 ACAATATGTTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1233
Db 39969 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40028
QY 1234 TTAAGGAAGTTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1293
Db 40029 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40087
QY 1294 GAGACTAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1353
Db 40088 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40147
QY 1354 TGGACCGTATGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1413
Db 40148 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40207
QY 1414 AAACAGAAAAAAGAAACATTTTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1473
Db 40208 GTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40267
QY 1474 ATCTTTGGTGAAGTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1533
Db 40268 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40327
QY 1534 TTGCTTTCTTTAATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1593
```

D	40386	TATATTATAAATTAATATATTTTAA--TTATTTTATTAATATTTAATTAATTTTAA	40386
Oy	1594	CTTGACATTCACAATTTTTATATAGATGATGTGATTAAT--TCCAAATACATAC	1651
D	40387	TTTTTAACTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40446
Oy	1652	TTTGGATGCTTTAAACTTAAATCTGTTTCTCCTACGGTAAATATTAATCATCGGTA	1711
D	40447	TTTTTAATTAATTAATTTTAATTTAATTTATTAATTAATTTAATTAATTTAATTA	40506
Oy	1712	AAAAAAGTTTGTCTTATTTTTCGCAGTAGCAAGATTAACCTAATGACTTTAATTTT	1771
D	40507	AATATTTATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40566
Oy	1772	TGAAAATGTAACCTTTTACT----CATGATTAATTAACCGATGTTTTTGTGCCATA	1827
D	40567	ATTTTAAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40626
Oy	1828	TGACAGCCCTTCACAAGTGA-----TAGCAATTTTTCGCAAAATTAATTAATAG	1880
D	40627	TTAATTTAAATTAATTAATTTAATTTAATTAATTAATTAATTAATTAATTAATTTTA	40686
Oy	1881	AATTCATGCTACATCATAGAGAAGAACGCTGATTAATTAATTTAATTTAATTAAGCAA	1940
D	40687	ATTAATTAATTAATTTAATTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTA	40746
Oy	1941	AATTTTGGAAAAAGTTATATTTCTACAAATTAATTTAATTAATTAATTAATTAATTA	2000
D	40747	ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40806
Oy	2001	TTTCCATGCTTAAATTAATTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTA	2060
D	40807	TATATTTAATTTAATTTAATTAATTAATTAATTTAATTTAATTTAATTTAATTAATTAAT	40864
Oy	2061	AATAGTGGTGAATTCAAATATCTCCATTAATTTTAAATTTTGAATCTCACAAATTAATTA	2120
D	40865	TATATTTAATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAAT	40924
Oy	2121	TTTACTCAATACAAATGCAATAGAAAAGTTCCAAAAAATTTGTTACAGAAAGCTCCAA	2180
D	40925	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAATTTAATTAAT	40984
Oy	2181	ATTTTTTTTTTTTATGAGAAAGAAATACATACATGAGAAAAGCAATTTGTTGGAATGGA	2240
D	40985	ATATTTTATATGTTAATTAATTAATTAATTTAATTTAATTTAATTAATTAATTAATTAATTA	41044
Oy	2241	AGTACTAATATACATTAAGCAAAATTTTAAAAAATTAATTAATTAAGCTATACGCGCTCAAAAT	2300
D	41045	TTTAAATTTAATTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT	41104
Oy	2301	ATGTTATCTAGTAGTGTAATTAATTAATGCAATGTCGATTCAGAAATGGGCAACAAATG	2360
D	41105	ATTATATATATATTTATACACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATG	41163
Oy	2361	AAAAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2407
D	41164	TATATTTAATTAATTTAATGAATTTAATTTAATTTGATTTAATTAATTA 41210	
<hr/>			
RESULT 11.			
AC120883/c 13877 bp DNA linear HTG 06-AUG-2007			
LOCUS Homo sapiens chromosome 18 clone CTD-2146H24 map 18, WORKING DRAFT			
DEFINITION SEQUENCE, 4 ordered pieces.			
AC120883 AC120883.2 GI:22123189			
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE 1 (bases 1 to 13877)			

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren, B., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 18, clone CTD-2146H24
Unpublished

TITLE	JOURNAL	REFERENCES	AUTHORS
1. The Effect of Temperature on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodate	Journal of Chemical Education	1. Smith, J. D. <i>J. Chem. Educ.</i> 1965, 42, 123.	Smith, J. D.
2. The Effect of Concentration on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodate	Journal of Chemical Education	2. Jones, A. B. <i>J. Chem. Educ.</i> 1968, 45, 456.	Jones, A. B.
3. The Effect of pH on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodate	Journal of Chemical Education	3. Brown, C. D. <i>J. Chem. Educ.</i> 1970, 47, 789.	Brown, C. D.
4. The Effect of Catalyst on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodate	Journal of Chemical Education	4. Green, E. F. <i>J. Chem. Educ.</i> 1972, 49, 1012.	Green, E. F.
5. The Effect of Surface Area on the Rate of Reaction of Hydrogen Peroxide with Potassium Iodate	Journal of Chemical Education	5. White, G. H. <i>J. Chem. Educ.</i> 1974, 51, 234.	White, G. H.

Birren, B., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 18, clone CTD-2146H24
 unpublished
 2 (bases 1 to 133877)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukanger, B., Brown, A., Camarata, J., Campolongo, A., Chang, J.,
 Chazaro, B., Chepelev, Y., Colangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., Dearlano, K., Desair, K., Diaz, J.S., Dodge, S.,
 Fark, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamet, A., Karatas, A., Kells, C., Lacroque, K., Lamzares, R.,
 Lander, T., Lehotzky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarty, M., McEwan, P., McKernan, K., Meldrum, J., Meneses, L.,
 Mihova, T., Mlenga, V., Murphy, J., Naylor, J., Nguyen, C., Nicoll, R.,
 Norby, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rossetti, M., Roy, A., Santos, R., Schauer, S., Schipbach, R.,
 Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tsefaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission

TITLE
JOURNAL
COMMENT

Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genomic Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 133877)
Barron, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camarate, J., Chang, J., Chazaro, B., Choepel, K., Collamore, A.,
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gird, S., Gisham, L., Grand-Platier, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Mayor, T., Matthews, C.,
McCarthy, M., Meldrum, J., Menais, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicot, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Rhunkhang, P., Pterre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schumpack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Strange-Thompson, N., Stojanovic, N., Talamas, J.,
Tefaye, S., Theodore, J., Tophan, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zalnoun, J.,
Zemбек, L., Zimmer, A. and Zody, M.

Submitted Submission
 Duplicate (06-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 6, 2002 this sequence version replaced g1:20531967.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/M/RepeatMasker.html>
 Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L26769
Center clone name: 2146_H_24

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 132654 bases at least Q40
Consensus quality: 132863 bases at least Q30

Consensus quality: 133065 bases at least Q20

Insert size: 133000; agarose-fp
Insert size: 133577; sum-of-contigs

Quality coverage: 17.1 in Q20 bases; agarose-fp

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db	Sequence	Position
Db 94820	ATTATATATTTTTCATATATTAATTTATATATATTTATATATATATATATATATATATATTA	94761
Qy 1727	TATTTCCGGATGCATGAAGGATAAAGCTTAATGACTTTATATTTTGGAAATGTAAACCT	1786
Db 94760	AATTTATTTATTTATTA	94701
Qy 1787	TTTACTCATAGATTAATTAACGTATGTTTCTTTCGCATTAAGACAGCCTCAGCAAGT	1846
Db 94700	TATATATATTTATTT	94641
Qy 1847	GATAGTCATTTTCTGCAATATATTAATTAAGAAATTCAGAAATTCATATATAGAGA	1906
Db 94640	ATTAT	94584
Qy 1907	AACAGCTGAGTATTCATTTTATTTTAAAGACAAAATTTTGCAAAAGTTTAAATTTCT	1966
Db 94583	TAAAT	94524
Qy 1967	AACAT	2026
Db 94523	TAAAT	94464
Qy 2027	TTTAT	2086
Db 94463	ATTAAT	94404
Qy 2087	ATTAAT	2146
Db 94403	ATTAT	94344
Qy 2147	TTCCAAAAAAATTTTGTGAACAGAACTTCCAAATTTTTTTTTTTTATGAGAACAGAA	2206
Db 94343	TAT	94284
Qy 2207	TAAACAGATAGAAAATCTATTTTGTGCAATGGAAGTATATATATATATATATATATATAT	2266
Db 94283	ATTAAT	94224
Qy 2267	TAAAAAATTTATTAAGCTATACGCGCT--CAAGATAGTTATCTAGTGAATTA	2323
Db 94223	TATTTAT	94164
Qy 2324	ATTAATGATGGGCGCATTCAGAAATGGGACACAATGAAAACGGAATTTAAATATATACT	2383
Db 94163	AAT	94104
Qy 2384	TTAAAT	2425
Db 94103	ATTAT	94062

RESULT 12

PFMAL3P5/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

86827 bp

MAUZP5, complete sequence.

AL010206 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162

AL010206 AL001210 AL139179

AL034556.3 GI:7711064

HNG: centromere; CTRP proteol: initiation factor E4;

Setine/threonine protein phosphatase.

Plasmodium falciparum 3D7.

Plasmodium falciparum 3D7.

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

1 (bases 1 to 86827)

Boman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Feltham, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jags, K., Jassal, B., Kyes, S., McLean, J., Moun, S., Mungall, K., Murphy, J., Oliver, K., Ouali, M.A., Rajandream, M.-A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Whitehead, S., Woodward, J.R., Newbold, C. and

Brettell, B.G.

TITLE	FEATURES
JOURNAL MEDLINE PUBMED AUTHORS	The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum Nature 400 (6744), 532-538 (1999) 99376085 10448855 2 (bases 1 to 86827) Bowman,S., Skelton,J., Churcher,C., Lawson,D., Quall,M. and Barrell,B.
JOURNAL REFERENCE AUTHORS TITLE JOURNAL	Unpublished 3 (bases 1 to 86827) Lawson,D., Bowman,S. and Barrell,B. Direct Submission Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium. The Sanger Centre, Wellcome Trust Genome Camps, Hinxton, Cambridge CB10 1SA, UK
COMMENT	On or before May 14, 2001 this sequence version replaced gi:2982535, gi:2982536, gi:2894454, gi:2982554, gi:2982562, gi:2894489, gi:2982572, gi:2982574, gi:4493931. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum .
FEATURES	Location/Qualifiers
SOURCE	1. 86827
	/organism="Plasmodium falciparum 3D7"
	/strain="3D7"
	/db_xref="taxon:36329"
	/chromosome="3"
gene	/clone="MAL3P5"
	join(322..603,826..1023,1197..1301,1458..2942)
	/gene="PFC0575w, MAL3P5.1"
	join(322..603,826..1023,1197..1301,1458..2942)
CDS	/gene="PFC0575w, MAL3P5.1"
	/note="PFC0575w (MAL3P5.1), Hypothetical protein, len: 689 aa, possible signal sequence, revised: added new exon 2"
	/codon_start=1
	/product="Hypothetical protein, PFC0575w"
	/protein_id="CA38969.2"
	/db_xref="GI:7711065"
	/db_xref="SPTREMBL:097258"
	/translation="MYLKNVYIYISCFILFDLCFSFHLKMKYKKNMNMKSVTEFL RSQIYKRRKRSRIRKKNVSKKQKPLFPEMLKGFSELGFWRNQYDQYIDYVTS NINNLRIQVTHKHKSEFTEKRIQILILCVFSKIDFKILNLSYIKHFSQNSTI VHSILNDSIKVEKKEKDAENYVLAHLHLEIDNTLFSMHHIDFESKQVCEICD IKKKRKKNLSTIYINFICTILYFYCKMLILKLYISHLCTIPFFPFCFLCYLIER IYECVGDILRKKILERTNLCEKKIKLFHKHDAIKKEIMKKDDLYFNHYDLKLC FPMKLNTERNNKNIIRSNYNININDSIDDMNPNIDVNIINNIISLEKIKQEFNP DDENLKELKDTQEPOLFENYNIINIKYEEDQPLVNIINNNIINPMKMKHKIK DYVNDNDQDYDEKEEDVLYQKNIDYIKYKTIQNKNSLEEFKQFEOADIEPQNFIS NVLNDGGRYKSNDEKTSREHKTKKNTKIGYDTLIDQNMENFTIKKINDNISD ISNNEIYIKIKLNQSDSDINLTSDLIYRLKTKYMTYQKIEYLFKFIQYDILNQ YPIKKEKTVLDLNLGTGKIVMSPDVNSLFEKTKIDSIPNEKDKNMQKNSKNY IMFNWA"
misc_feature	604..609
	/gene="PFC0575w, MAL3P5.1"
	/note="predicted splice donor sequence for exon 1 of PFC0575w (revised)"
misc_feature	816..825
	/gene="PFC0575w, MAL3P5.1"
	/note="predicted splice acceptor sequence for exon 2 (revised of PFC0575w)"
misc_feature	1024..1029
	/gene="PFC0575w, MAL3P5.1"
	/note="predicted splice donor sequence for exon 2 of PFC0575w (revised)"
gene	complement(3354..6644)
	/gene="PFC0580c, MAL3P5.2"
	complement(3354..6644)
	/gene="PFC0580c, MAL3P5.2"
CDS	/note="PFC0580c (MAL3P5.2), Hypothetical protein, len: 1097 aa, possible signal sequence, predicted using hexon"
	/codon_start=1
	/product="Hypothetical protein, PFC0580c"
	/protein_id="CA38968.1"

misc_feature	/note="potential splice donor sequence for exon 5 of PFC0581w"	8559..8567	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 6 of PFC0581w"	8642..8647	/gene="PFC0581w"	/note="potential splice donor sequence for exon 6 of PFC0581w"	8704..8712	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 7 of PFC0581w"	8813..8818	/gene="PFC0581w"	/note="potential splice donor sequence for exon 7 of PFC0581w"	8913..8923	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 8 of PFC0581w"	9007..9012	/gene="PFC0581w"	/note="potential splice donor sequence for exon 8 of PFC0581w"	9112..9121	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 9 of PFC0581w"	9259..9264	/gene="PFC0581w"	/note="potential splice donor sequence for exon 9 of PFC0581w"	9361..9368	misc_feature
--------------	--	------------	------------------	---	------------	------------------	--	------------	------------------	---	------------	------------------	--	------------	------------------	---	------------	------------------	--	------------	------------------	---	------------	------------------	--	------------	--------------

Query Match	Best Local Similarity	Score	DB 3:	Length
Matches 1023;	Conservative	0;	Mismatches 1266;	Indels 38; Gaps 10;
QY 108	ATAATTAATAATTAATAATTCATTTAAATATTTACAGAGTAATTAATATCTTTACATTGT	167		
DB 41351	ATTAATTAATTTAGCAATTTATTTAAATTAATAATTTAAATTCAGAAAATTAATTAATATA	41292		
QY 168	ATTGTTAAACAAAATATCTATCTTGGTATATGAGAAATATGAGATTGGAAATTTATA	227		
DB 41291	AAATAAAATATACACATCAATATAAAATATATCATTTATTAATAATTAATTAATTAATTA	41232		
QY 228	ATTAATTAAGGAATATTCATTCATCTTGGTATACACAGTTAAAGTTTGTGTTTC	287		
DB 41231	ATTTTATTAATAATTTAAATTAATAATTAATAAAATCAATTAATAATTAATTCATTAATTAAT	41172		
QY 288	TTTTGTTATATGTAATATGATGAATCAAAAGAGATTTGAGTGTAACATATTTTC	347		
DB 41171	TAAATATACATTAATTAATTTAAATTAATAATTAATTAAT-----AATTAAGGT	41121		
QY 348	GTTATGACCCCAAAAAAAAAAAAAAAAACAACAACAAACCCCCCGCATATAGTTT	407		
DB 41120	ATAATAATATATATTAATTAATAATTAATAATTTAAATTTAAATTAATTAATTAATTAATTAAT	41061		
QY 408	TTTGTTCTGATTTAGGTTTATTTGATCAATATACATGCATCTTTCTTTGATTAAGTATG	467		
DB 41060	TCATTAATAATTAATTAATTAATTAATTAATAATTAATTAATTAATAATTAATAATTAATTAAT	41001		
QY 468	AGATTTTCTTACCAATTAATAATTTGCATCAATCTCTGATT---ATTAATTTAAAT	524		
DB 41000	TAACTATATAATATATATTAATAATTAATAATTAATTAATTAATTAATTAATTAATTAAT	40941		
QY 525	ACGATGTGGAATATCCGTTTATCGATCACTCCAAATCATGATTATGATCTGTGCTATATC	584		
DB 40940	TATTAATAATTAATAATAATAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40881		
QY 585	CAGCAAAATTAATTAACAAGATTTGAGAAAAAACCGAAATATACAAAAGGGAAGGAGTAG	644		

[illegible]

Dd	39811	ATAAATATATAATTAATTTATTT-----TATATATATATATATATTTATTTATTAAC	39760
Qy	1724	TCTTATTTTCGGGATGATGAGGATTAACCTTAATGACTTAAATTTTTTGAAAAATGAAC	1783
Dd	39759	GGTATTTTAAATTCATTATTAATATATATATATATTAATTAATTAATTAATTAATTTAT	39700
Qy	1784	CCCTTACTGATAGATTAATTAATCCGATGCTTTTGTCGCAATAGACAGCCCTACACAC	1843
Dd	39699	TATTAATTTTATTAATTAATTAACCTTTATTTAT--TTAATATTTAATTTAATTAATATA	39641
Qy	1844	TGTGATGTCATATTTTTCTCGCAATATTAATTAATGAAATCAATGCTACTATCAATAGA	1903
Dd	39640	TATATATATATATATTTATTTGAATTAATTAATTTATTTATTTATTTATTTATATATAT	39581
Qy	1904	AGAAACAGCTGAGATTAACATTTTAATTTAAAGACAAAATTTTGAAAAATGTATTAAT	1963
Dd	39580	ATATATTAATTAATTTATTTTAAATTTATTTATTTATTTATTTATTAATTAATTAATATA	39521
Qy	1964	TCTACATATTTATTAATAATATGATGCCATATATGATTTCCATGCTCTTAAATATTT	2023
Dd	39520	TTAATTTTATTTTAAATTAATTAATTTATTTTATTTTAAATTAATATATATTTAAT	39461
Qy	2024	TTTTTATATTTAGTATTAATTAACATTAATTAAGAACATATAGTGTGTAATCAATATAC	2083
Dd	39460	TATTTTATTTTAAATTAATTAATATATATTTATTTTATTTTAAATTAATTAATATA	39401
Qy	2084	TCCATTAAATTTTGTGAATCTACAAATTAATTAATTTAGTCAATACATACGATAGA	2143
Dd	39400	TTAATTTTATTTTAAATTAATTAATTAATTAATTTATTTTATTTTAAATTAATATA	39341
Qy	2144	AAAGTCCAAAAAAATTTTGTTAACAGAACTCCAAATTTTTTTTTTTATATGACACAG	2203
Dd	39340	TATATATTAATTTATTTTATTTTAAATTAATTAATTAATTAATTAATTTATTTTAT	39281
Qy	2204	AAATA---ACAGATAGAAACATATTTGTGCGAATGGAAGTACTAGTAATATCATTAAAC	2260
Dd	39280	AAATATATATTAATTTATTTTATTTTAAATTAATTAATTAATTAATTTTATTTATTTT	39221
Qy	2261	AAATTTTAAAAAATTAATTAAGCCATACAGCCCTCAAGATATGTAATCAGTAGGTGTA	2320
Dd	39220	AAATATATATATTTTATTTTAAATTAATTAATTAATTAATTTATTTATTTGTTAT	39161
Qy	2321	TTAA--TAATGATGCTGCGATTCAGAAATTTGGACACACATGAAAAACGAAATTAATAAT	2378
Dd	39160	TTTATTTTATTAATTAATTAATTAATTAATTTATTTTATTAATTAATTAATTTATTAAT	39101
Qy	2379	TAACTTAAATTAATTAATTAATTTAGTAATAATGCTTTTCGACAT	2425
Dd	39100	TTAATATTAATTAATTAATTAATTAATTAATTTTATTTATTTT	39054

RESULT 13					
AC105425/c					
LOCUS	AC105425	205130 bp	DNA	linear	HTG 15-JAN-2002
DEFINITION	Homo sapiens chromosome 7 clone RP11-776N17, WORKING DRAFT				
	SEQUENCE, 12 unordered pieces.				
ACCESSION	AC105425				
VERSION	AC105425.2	GI:18151067			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens.				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 205130)				
AUTHORS	Waterston, R.H.				
TITLE	The sequence of Homo sapiens clone				
JOURNAL	unpublished				
REFERENCE	2 (bases 1 to 205130)				
AUTHORS	Waterston, R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-JUN-2002) Genome Sequencing Center, Washington				
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
	MO 63108, USA				

OY	903	ACTATATTTTTCGCGGGGATATATGATATTAACAAGATATATCAACAAAACATTCCTCGGGAC	962
Db	78934	ATCATTATATATATTTGTTATATATATGTAACATATATATATATATTTTGTTATATATGTAAC	78875
OY	963	CATTTCGAATAAACCCTTTTTCGAAACATTTAGCGGACCTGCACCTGCACCTTAAATAGC	1022
Db	78874	TAT	78815
OY	1023	ATTTCACGCGCTACTAGTGGATTACTAGATTAAGCATTAAGGACCCGTTCAAGCTA	1082
Db	78814	AATGATATTTTATATATATTAACATAAATTTATATATATATATATATATATATATATATATAT	78755
OY	1083	TTTATACAAAGTGTCAAACTGAAATATAGCTTGAATCCCTTGGAAAATTTTGGAAATTTAC	1142
Db	78754	TAT	78655
OY	1143	GCTGTATATGTAATAATATAGATTTACTGGTAAACAAATATGTTAAATCAATTTAGTGGTCAAC	1202
Db	78694	TTAT	78636
OY	1203	ATATACATATATCCCTTACAGAAAAACAACCTTAAGAGAGTTAAACATATCCATATATGG	1262
Db	78635	TAT	78576
OY	1263	GTATGCTATACCTTCCAGGTATGCTATATAGACTAAGAGCTAAAGAAATGTTATGATGTGCA	1322
Db	78575	ATTAAT	78516
OY	1323	TAAATGAAATTCACACGGCGTGTAATTAATGAGACCGTATGTCAGTACCTCAAT	1382
Db	78515	TAT	78466
OY	1383	ATCATTTCTGGTGGTCAACAATTAACAAAAACAAGAAAAAAGAAACGATTTTCT	1442
Db	78465	TTAT	78406
OY	1443	TGGATTCATTCACATGATCTTAATAAGCATATGCTTTGGGTTACAGTTTGGAAATCCGC	1502
Db	78405	TAT	78346
OY	1503	TACAAGCGTGAACCACTGCAACTTAATTAATTTGCTTTCTTAATGCATCTTTAAACATAT	1562
Db	78345	TACAT	78287
OY	1563	TTATTTGTATGTTGGAAATTAATAGACGGACTGTGAACATTACATATTTATATAGAT	1622
Db	78286	AT	78227
OY	1623	ACTAGTATGTGATTTATCCAAATACATACCTTGGATTTAACTTAATCTGTTCTTC	1682
Db	78226	CATATACCTAT	78167
OY	1683	CTACGATTAATAATATTAATCATCGAGTAAAAAAGTTTGTCTTATTTTTCGCGATGCAT	1742
Db	78166	TAT	78107
OY	1743	GAGGATTAACCTATATGACTTTAATTTTTCGAAAATGTAACCCCTTTTACTCATGATTA	1802
Db	78106	ATTATATATAGAT	78048
OY	1803	TTACGCTATGTTTTGTTGCCATATGACAGCGCTTACACGCTGATAGTCACTATTTTC	1862
Db	78047	ATATATCATATATTTATTAACAGCTTATATATATATATATATATATATATATATATATATATTA	77988
OY	1863	TGCACAAATTTTAAATAGCAATTCATGCTACTATCAATAGAGAAGAACGCTGAGATATAC	1922
Db	77987	TAT	77928
OY	1923	ATTTTAAATTTAAAGCAAAATTTTGA AAAATGTTATATATTTCTAACAATATATTTAAAA	1982
Db	77927	ATTAT	77868
OY	1983	TATGATGCCCTAATATGATTTCCCTATGTTCTTAAAAATTTTTTTTTTTTAAATTTACTTATA	2042

[illegible]

```

RESULT 14
DMU11584
LOCUS          4601 bp      DNA      linear      INV 23-JUL-1994
DEFINITION    DMU11584
ACCESSION     U11584
VERSION       U11584.1 GI:508826
KEYWORDS      mitochondrial DNA; A+T region; tandem repeats.
SOURCE        fruit fly.
ORGANISM      Mitochondrion Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
               Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
               Ephydroidea; Drosophilidae; Drosophila.
REFERENCE     1 (bases 1 to 4601)
               Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
               Sequence, organization, and evolution of the A+T region of
               Drosophila melanogaster mitochondrial DNA
               Mol. Biol. Evol. 11 (3), 523-538 (1994)
JOURNAL       94285822
MEDLINE       8015445
PUBMED        2 (bases 1 to 4601)
REFERENCE     Kaguni,L.S.
               Direct Submission
               Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D. Dept. of
               Biochemistry, Michigan State University, East Lansing, MI,
               48824-1318, USA
TITLE         location/Qualifiers
JOURNAL
FEATURES
               source
               1..4601
               location/Qualifiers
               /organism="Drosophila melanogaster"
               /organelle="mitochondrion"
               /strain="Oregon-R"
               /db_xref="taxon:7227"
               /dev_stage="embryo"
               650..1022
               /note="repeat I-A"
               /rpt_type=tandem
               1023..1360
               /note="repeat I-B1"
               /rpt_type=tandem
               1361..1705
               /note="repeat I-C-A"
               /rpt_type=tandem

```


Db 3941 ATTATGAAATATTTATATATATATATAT-ATATATAGAAAAATTAATTTAATTA 3999
 QY 1787 TTTACTCATAGATTAATTAACGGATGTTTGTTCCTAATGACAGCCCTACAACTGT 1846
 Db 4000 TTTAATATATATTTTAAATTTCTTAATGTAATTTTATATATATATATATATAT 4059
 QY 1847 GATAGTCAATTTTTCGCAATATTAATTAATGCAATTCATGCTATCATGCAAGA 1906
 Db 4060 AATAAATCAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4119
 QY 1907 AACAGCTGAGTATTTATTTATTAATTAAGCAAAATTTTGAAGAAATGTTAATTTCT 1966
 Db 4120 AATATATATTTATTTATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAT 4179
 QY 1967 AACATATTTATTAATTAATGATGCTATTAATGATTTCTATGTTCTTAATATTTT 2026
 Db 4180 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4239
 QY 2027 TTTATATTTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2086
 Db 4240 TTAATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4299
 QY 2087 ATTAATATTTTGAATCTCAAAATTTATTAATTTATGCTAATTAACATGATGAAG 2146
 Db 4300 ATATATATTTATTAATTAATTAATTTCTCATTTATTTATTAATTTATTAATTA 4359
 QY 2147 TTCCAAAAAATTTTGTAAACAGAACTCCAAATTTTATTTTATGCAACAGAA 2206
 Db 4360 TAAATATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4419
 QY 2207 TAACGATGAGAACTATTTTGTGTGAATGGAAGTATTAATTAATTAATTAATTA 2266
 Db 4420 TATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4479
 QY 2267 TAAATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2326
 Db 4480 AAAAAATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4539
 QY 2327 ATGATGCTGCGATTCAGAAATGGAACAATGAAACGAAATTAATTAATTAAT 2386
 Db 4540 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4599

RESULT 15
 LOCUS DM037541 19517 bp DNA circular INV 02-MAR-2001
 DEFINITION Drosophila melanogaster complete mitochondrial genome.
 ACCESSION U37541
 VERSION U37541.1 GI:1166529
 KEYWORDS
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 12511 to 12682)
 Clary,D.O., Goddard,J.M., Martin,S.C., Faumon,C.M. and
 Wolstenholme,D.R.
 Drosophila mitochondrial DNA: a novel gene order
 Nucleic Acids Res. 10 (21), 6619-6637 (1982)
 REFERENCE 2 (bases 5269 to 5695)
 Clary,D.O., Wahlthner,J.A. and Wolstenholme,D.R.
 Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
 flanking sequences and comparisons to mammalian mitochondrial tRNA
 genes
 Nucleic Acids Res. 11 (8), 2411-2425 (1983)
 JOURNAL MEDLINE 8320794
 PUBMED 6304652
 REFERENCE 3 (bases 404 to 5272)
 de Bruijn,M.H.
 Drosophila melanogaster mitochondrial DNA, a novel organization and

JOURNAL MEDLINE 83245048
 PUBMED 6408489
 REFERENCE 4 (bases 804 to 1778)
 Satta,Y., Ishiwa,H. and Chigusa,S.I.
 Analysis of nucleotide substitutions of mitochondrial DNAs in
 Drosophila melanogaster and its sibling species
 Mol. Biol. Evol. 4 (6), 638-650 (1987)
 JOURNAL MEDLINE 88174373
 PUBMED 2832697
 REFERENCE 5 (bases 5268 to 13619)
 Garesse,R.
 Drosophila melanogaster mitochondrial DNA: gene organization and
 evolutionary considerations
 Genetics 118 (4), 649-663 (1988)
 JOURNAL MEDLINE 88212147
 PUBMED 3130291
 REFERENCE 6 (bases 441 to 2967)
 Satta,Y. and Takahata,N.
 Evolution of Drosophila mitochondrial DNA and the history of the
 melanogaster subgroup
 Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
 JOURNAL MEDLINE 91088557
 PUBMED 2124697
 REFERENCE 7 (bases 14215 to 14512)
 Ballard,J.W., Olsen,G.J., Falch,D.P., Odgers,W.A., Rowell,D.M. and
 Atkinson,P.W.
 Evidence from 12S ribosomal RNA sequences that onychophorans are
 modified arthropods
 Science 258 (5086), 1345-1348 (1992)
 JOURNAL MEDLINE 93088057
 PUBMED 1455227
 REFERENCE 8 (bases 14917 to 19517)
 Lewis,D.L., Farr,C.L., Farguhar,A.L. and Kaguni,L.S.
 Sequence, organization, and evolution of the A+T region of
 Drosophila melanogaster mitochondrial DNA
 Mol. Biol. Evol. 11 (3), 523-538 (1994)
 JOURNAL MEDLINE 94285822
 PUBMED 8015445
 REFERENCE 9 (bases 1 to 408; 13319 to 19517)
 Lewis,D.L., Farr,C.L. and Kaguni,L.S.
 Drosophila melanogaster mitochondrial DNA: completion of the
 nucleotide sequence and evolutionary comparisons
 Insect Mol. Biol. 4 (4), 263-278 (1995)
 JOURNAL MEDLINE 96423163
 PUBMED 8825764
 REFERENCE 10 (bases 1 to 19517)
 Lewis,D.L., Farr,C.L. and Kaguni,L.S.
 Direct Submission
 Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
 Michigan State University, East Lansing, MI 48824-1319, USA
 location/Qualifiers
 1. 19517
 /organism="Drosophila melanogaster"
 /organelle="mitochondrion"
 /db_xref="taxon:7227"
 /note="derived from new and previously submitted
 sequences; sequence is a composite containing sequences
 obtained from different Drosophila melanogaster strains"
 1..65
 /product="tRNA-ile"
 complement(97..165)
 /product="tRNA-gln"
 171..239
 /product="tRNA-phe"
 240..1265
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 2"
 /protein_id="AAC47810.1"
 /db_xref="GI:1166530"
 /translation="MFNNSKILFITTIIGTLITVTSNGWLGAMGLEINLISFIPL
 genetic code
 Nature 304 (5923), 234-241 (1983)
 83245048
 6408489
 4 (bases 804 to 1778)
 Satta,Y., Ishiwa,H. and Chigusa,S.I.
 Analysis of nucleotide substitutions of mitochondrial DNAs in
 Drosophila melanogaster and its sibling species
 Mol. Biol. Evol. 4 (6), 638-650 (1987)
 88174373
 2832697
 5 (bases 5268 to 13619)
 Garesse,R.
 Drosophila melanogaster mitochondrial DNA: gene organization and
 evolutionary considerations
 Genetics 118 (4), 649-663 (1988)
 88212147
 3130291
 6 (bases 441 to 2967)
 Satta,Y. and Takahata,N.
 Evolution of Drosophila mitochondrial DNA and the history of the
 melanogaster subgroup
 Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
 91088557
 2124697
 7 (bases 14215 to 14512)
 Ballard,J.W., Olsen,G.J., Falch,D.P., Odgers,W.A., Rowell,D.M. and
 Atkinson,P.W.
 Evidence from 12S ribosomal RNA sequences that onychophorans are
 modified arthropods
 Science 258 (5086), 1345-1348 (1992)
 93088057
 1455227
 8 (bases 14917 to 19517)
 Lewis,D.L., Farr,C.L., Farguhar,A.L. and Kaguni,L.S.
 Sequence, organization, and evolution of the A+T region of
 Drosophila melanogaster mitochondrial DNA
 Mol. Biol. Evol. 11 (3), 523-538 (1994)
 94285822
 8015445
 9 (bases 1 to 408; 13319 to 19517)
 Lewis,D.L., Farr,C.L. and Kaguni,L.S.
 Drosophila melanogaster mitochondrial DNA: completion of the
 nucleotide sequence and evolutionary comparisons
 Insect Mol. Biol. 4 (4), 263-278 (1995)
 96423163
 8825764
 10 (bases 1 to 19517)
 Lewis,D.L., Farr,C.L. and Kaguni,L.S.
 Direct Submission
 Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
 Michigan State University, East Lansing, MI 48824-1319, USA
 location/Qualifiers
 1. 19517
 /organism="Drosophila melanogaster"
 /organelle="mitochondrion"
 /db_xref="taxon:7227"
 /note="derived from new and previously submitted
 sequences; sequence is a composite containing sequences
 obtained from different Drosophila melanogaster strains"
 1..65
 /product="tRNA-ile"
 complement(97..165)
 /product="tRNA-gln"
 171..239
 /product="tRNA-phe"
 240..1265
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 2"
 /protein_id="AAC47810.1"
 /db_xref="GI:1166530"
 /translation="MFNNSKILFITTIIGTLITVTSNGWLGAMGLEINLISFIPL

Db 1777 ATATATCTAATTTTAAATTAATTAATTTTAAATTTTAAATTTAGATATATTTATA 17836
Qy 724 AGCTTCTCTCAATCAATATGACCTATTTGGATTTTCGATATTCGTTAAATTT 783
Db 17837 AAATTTATATCTCAATATTTATTTATTTAATTTAATTTAATTTAATTTAATTT 17896
Qy 784 GCGATAGATTTGTGAAA--AATATTTTGTGTAGCTGATCTCAATTTATGT-TCGA 840
Db 17897 TAATTAATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTA 17956
Qy 841 GGTATTTGCAATATCTGTTTAAAGCATATTTGCTTTCTTTGTTGTTGTTCTCT 900
Db 17957 TATATATGAAAAATTTAATTTTAAATTTAATTTAATTTAATTTTAAATTTTCTTA 18016
Qy 901 TAACATATATTTATCGCGGATATAT---GATTAACATGATATATACCAAAACATTTGT 955
Db 18017 AATGATATTTATTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTAAC 18076
Qy 956 CTGGGACCATTTGTAATAACCTTTTCTCAACATTTACGGGACCTGACCTGACCCCTTA 1015
Db 18077 AAAAAATTTTATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTTCAATTTT 18136
Qy 1016 AAATACGATTTTACAGCGTCACTAGTTAGATTTAGCATTAAGCATTAAGACCCGTT 1075
Db 18137 TTAATAAATTTTAAAAAATTTATTTTAAAAAATTTAATTTAATTTAATTTAATTTA 18196
Qy 1076 CA----AGCATTTTATACAAAGTTACAAACTGAATATAGCTTGAATCTTTAGAAAT 1130
Db 18197 AATTAATGATTTTATTTATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 18256
Qy 1131 TTTGAATTTACCGTTGTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 1190
Db 18257 AAATAAAAATTTTAAAAATTTTAAAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 18316
Qy 1191 TTATGTCACAACTATTAATTTCTTACAGAAAAACAACTTAAGAGAGTTAACT 1250
Db 18317 ATTTATTTATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 18376
Qy 1231 ATTCATATATGGGTATGCTATACCTTTCACTGATCTATAGACTTAAGAAATAGTT 1310
Db 18377 ATTTTAAATTTTATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 18436
Qy 1311 ATGATGTCGATTAATGAAATTCACAGCGGTGTAATATTT--ATGGACCGCATGTT 1367
Db 18437 AAATATTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 18496
Qy 1368 ACATCACTGCAATATATCTTCTGTTGTCACATTAACAAAAACAAAAACAAAAA 1427
Db 18497 AAAAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18556
Qy 1428 GAAAAAGATTTTCTGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1487
Db 18557 ATTAATTTTATTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 18616
Qy 1488 AGTTTCGAGTCCTCTACAAGGCTGTAACCATCTGCACTATTAATGCTTCT-TTAA 1546
Db 18617 AAAAAATTTTATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 18676
Qy 1547 TGCATCTTTAATTTATTTATTTAGTTGAATTTAATTAAGAGCACTGTAACATTAC 1606
Db 18677 TATATATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 18736
Qy 1607 AATATTTTATTTAGATTTAGTATGATTTTCCAAATATACATCTTTGATGTTTAAAC 1666
Db 18737 AATTTAAAAATGATATTTAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 18796
Qy 1667 TTAATCTGTTCTCTCCACGTAATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1726
Db 18797 TTTAATTTATTAATTAATTAATTAATTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTA 18856
Qy 1727 TATTTTCCGATGATGAAGTAACCTTAATGATTTAATTTTGAATGTAACCTT 1786
Db 18857 ATTAATGATATTTTATTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 18915

Qy 1787 TTTACTATAGATTTAATTTACCGTATGTTTGTGCCAATATGACGCGCTACAACTGT 1846
Db 18916 TTTAATATTAATTTTAAAAATTTCTTAATGATATTTTATTAATTTAATTTAATTTAATTTAATTTA 18975
Qy 1847 GATAGCAATTTTCTCGAATATTAATTTAGAAATTCGAAGCTCTATCATTAAGAA 1906
Db 18976 AATTAATATTTTAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 19035
Qy 1907 AACAGCTGAGTATTTACATTTTAAATTTAAGACAAATTTTGAATAATGTTATATTTCT 1966
Db 19036 AATATTAATTTTATTTATTTTTCATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 19095
Qy 1967 AACATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 2026
Db 19096 TTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 19155
Qy 2027 TTTATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTATCTGC 2086
Db 19156 TTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAATTAATTA 19215
Qy 2087 ATTAATATTTTGAATTTACAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 2146
Db 19216 ATATATTTTAAAAATTTATTTATTTCTCATATTTATTTATTTAATTTAATTTAATTTAATTTA 19275
Qy 2147 TTTCAAAAAATTTTGTAAACAGAACTTCCAAATTTTAAAAATTTTAAAAATTTTAAAAATTTTAAAA 2206
Db 19276 TAAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 19335
Qy 2207 TAACAGATGAAGAACTATTTGTTGGAATGGAAGTAACTAATTTAATTTAATTTAATTTAATTTA 2266
Db 19336 TATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 19395
Qy 2267 TAAAAAATTTATTAAGCCCTATCGCGCTCAAGATATGATGATGATGATGATGATGATGATGATGATG 2326
Db 19396 AAAAAATTTCTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 19455
Qy 2327 ATGATGTTGCAATTTGCAATTTGGGCAATGAAACGAAATTAATTAATTTAATTTAATTTAATTTA 2386
Db 19456 TTTTAAAAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 19515

Search completed: March 30, 2003, 08:19:09
Job time : 34395.6 secs

FT exon 3914..4066
 FT /*tag- h
 FT Intron 4067..4164
 FT /*tag- i
 FT exon 4165..4479
 FT /*tag- j
 FT Intron 4480..4631
 FT /*tag- k
 FT exon 4632..4724
 FT /*tag- l
 FT Intron 4725..4815
 FT /*tag- m
 FT exon 4816..4894
 FT /*tag- n
 FT Intron 4895..5000
 FT /*tag- o
 FT exon 5001..5110
 FT /*tag- p
 FT Intron 5111..5864
 FT /*tag- q
 FT exon 5865..6110
 FT /*tag- r
 FT 3'UTR 6011..6468
 FT /*tag- s
 PN WO20047715-A2.
 XX 17-AUG-2000.
 PD 11-FEB-2000; 2000WO-US03820.
 XX 11-FEB-1999; 99US-0119657.
 PR 11-FEB-1999; 99US-0119658.
 XX
 PA (ARIZ-) ARIZONA BOARD OF REGENTS.
 PI Aspiroz R, Choe S, Feldmann KA;
 DR WPI: 2000-549142/50.
 DR P-PSDB: AAB07921.
 XX
 PT New isolated dwf4 polynucleotide useful for altering the phenotype of
 XX plants, for diagnostic assays and in the production of antibodies -
 PS Claim 3; Fig 10A-G; 113pp; English.
 CC The present sequence encodes a DMF4 polypeptide. The polypeptide is a
 CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
 CC brassinosteroids. Specifically, it mediates multiple
 CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DMF4
 CC polynucleotide is used for altering the phenotype of a plant. DMF4
 CC plants display a dramatic reduction in the length of different organs,
 CC and this size reduction is attributable to a defect in cell elongation.
 CC The DMF4 polynucleotides and polypeptides can be used in diagnostic
 CC assays and to generate antibodies, which can be used to produce
 CC immunogenic compositions.
 XX
 SQ Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;
 Query Match 100.0%; Score 3202; DB 21; Length 6888;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 AAAATTCATTTAAATTTTACAGTAATTAATTAATCTTACATCTGTTATTAACAA 180
 Qy 181 AATATCTATCTTGGATATGAAAAATATGAGTTGGAATTTAATAATAAGGAA 240
 Db 181 AATATCTATCTTGGATATGAAAAATATGAGTTGGAATTTAATAATAAGGAA 240
 Qy 241 TAATGATTCATTTGGTTGGATTACACAGTTAAAGTTTGTGTTCTTTGTTATATGT 300
 Db 241 TAATGATTCATTTGGTTGGATTACACAGTTAAAGTTTGTGTTCTTTGTTATATGT 300
 Qy 301 ATATGAGTAATCAAAAAAGATTTGATTTGAAGTGTAAACATATTTGTTAGCCCCCA 360
 Db 301 ATATGAGTAATCAAAAAAGATTTGATTTGAAGTGTAAACATATTTGTTAGCCCCCA 360
 Qy 361 AAAAAAACAACAAACAAACCAACCCCGCCCGATAGTTTGTGTTGGTTGGATT 420
 Db 361 AAAAAAACAACAAACAAACCAACCCCGCCCGATAGTTTGTGTTGGTTGGATT 420
 Qy 421 AGTTTATTTGATCATTAATTTACATGATCTTTGTTGATTAATGAGATTTTCTTAC 480
 Db 421 AGTTTATTTGATCATTAATTTACATGATCTTTGTTGATTAATGAGATTTTCTTAC 480
 Qy 481 CAATTAATTTGCAATTTATATCTCTTGTATTTAATAATAATACAGATGGAATATCC 540
 Db 481 CAATTAATTTGCAATTTATATCTCTTGTATTTAATAATAATACAGATGGAATATCC 540
 Qy 541 GTTATGATCATCTCCATCATGATTTATGATCTTGTGTTATTCAGCAATTTATTAACA 600
 Db 541 GTTATGATCATCTCCATCATGATTTATGATCTTGTGTTATTCAGCAATTTATTAACA 600
 Qy 601 AGATTTGAGAAAAACCGAAAAATAGAAAAAGGAAAGAGTAGTACCATGAGATATG 660
 Db 601 AGATTTGAGAAAAACCGAAAAATAGAAAAAGGAAAGAGTAGTACCATGAGATATG 660
 Qy 661 TGAATTAATTAACAAAGAGATTAAGATGACAAACCAAAAGTTGAGATATGTCCT 720
 Db 661 TGAATTAATTAACAAAGAGATTAAGATGACAAACCAAAAGTTGAGATATGTCCT 720
 Qy 721 GCCAGCTTCTCTCAACATCATATCGACCTATTGATTTGATTCGATATTCGTTAA 780
 Db 721 GCCAGCTTCTCTCAACATCATATCGACCTATTGATTTGATTCGATATTCGTTAA 780
 Qy 781 TTTGGATTAACGATTTGAAAAATATTTTATTTGAGCTGATCAATATATATGTCCA 840
 Db 781 TTTGGATTAACGATTTGAAAAATATTTTATTTGAGCTGATCAATATATATGTCCA 840
 Qy 841 GGTATTTGATTAATCTTCTGTTAAAGCATTTTGTGTTTGTGTTGTTGTTCTCT 900
 Db 841 GGTATTTGATTAATCTTCTGTTAAAGCATTTTGTGTTTGTGTTGTTGTTCTCT 900
 Qy 901 TAATATATATTTATGCGGATATATGATTAACATGATATATACAAACAAATTTGCTGG 960
 Db 901 TAATATATATTTATGCGGATATATGATTAACATGATATATACAAACAAATTTGCTGG 960
 Qy 961 ACCATTTGAAATTAATTTTCTCAAACTTTACGGGACCTGAGCTCGACCTTAAATA 1020
 Db 961 ACCATTTGAAATTAATTTTCTCAAACTTTACGGGACCTGAGCTCGACCTTAAATA 1020
 Qy 1021 CGATTTACAGGCTACATGATTTGATTAACATTAAGCAATTAAGGACCGCTCAAGC 1080
 Db 1021 CGATTTACAGGCTACATGATTTGATTAACATTAAGCAATTAAGGACCGCTCAAGC 1080
 Qy 1081 TATTTATCAAAAGTTACAACTGAATATAGTTGAATCTTTAGAAAATTTTGGAAATTA 1140
 Db 1081 TATTTATCAAAAGTTACAACTGAATATAGTTGAATCTTTAGAAAATTTTGGAAATTA 1140
 Qy 1141 CCGGTTGTTATGTAATATATGATTTAGTGTAAACAAATATGTTAATCAATTAAGGTCA 1200
 Db 1141 CCGGTTGTTATGTAATATATGATTTAGTGTAAACAAATATGTTAATCAATTAAGGTCA 1200
 Qy 1201 ACATATACATATCTCTACAGAAAAACAACTTAAGGAAGTTTAACATATCCATATAT 1260
 Db 1201 ACATATACATATCTCTACAGAAAAACAACTTAAGGAAGTTTAACATATCCATATAT 1260

```

QY 1261 GGGTATGCTATACCTTTCACGTATGCTATCTAGAGACTTAAAGATTAATGTTATGATGTC 1320
    |||||
Db 1261 GGGTATGCTATACCTTTCACGTATGCTATCTAGAGACTTAAAGATTAATGTTATGATGTC 1320
QY 1321 GATTAATGAAATTCACACGGGTGTATATATATGAGACGCTATGTTAGATCAGTCGCA 1380
    |||||
Db 1321 GATTAATGAAATTCACACGGGTGTATATATATGAGACGCTATGTTAGATCAGTCGCA 1380
QY 1381 ATATCATCTCTGGTGTGTCACATTAATAAACAACAAAGAAAAAGAAAGATTTT 1440
    |||||
Db 1381 ATATCATCTCTGGTGTGTCACATTAATAAACAACAAAGAAAAAGAAAGATTTT 1440
QY 1441 CTGGAATTCATCAATGATCACTTAAATGATGATCTTTGGGTTCAGAGTTGGAAGTCC 1500
    |||||
Db 1441 CTGGAATTCATCAATGATCACTTAAATGATGATCTTTGGGTTCAGAGTTGGAAGTCC 1500
QY 1501 TCTACAAAGCGTGTACACCTGTCACATCTTAAATGCTTTTAAATGATCTTTAAAT 1560
    |||||
Db 1501 TCTACAAAGCGTGTACACCTGTCACATCTTAAATGCTTTTAAATGATCTTTAAAT 1560
QY 1561 ATTTATTTAGTGTGATTAATAAGAGCACTTGTACATTAATTAATTAATAG 1620
    |||||
Db 1561 ATTTATTTAGTGTGATTAATAAGAGCACTTGTACATTAATTAATTAATAG 1620
QY 1621 ATACTAGTATGTGATTAATCCAAATACATCTTGTGATTAATCTTTTCT 1680
    |||||
Db 1621 ATACTAGTATGTGATTAATCCAAATACATCTTGTGATTAATCTTTTCT 1680
QY 1681 TCTACAGGCTAAATTAATTAATCATCGAGTAAAAAGTTTGTCTTATTTGGCGATGC 1740
    |||||
Db 1681 TCTACAGGCTAAATTAATTAATCATCGAGTAAAAAGTTTGTCTTATTTGGCGATGC 1740
QY 1741 ATGAAGATTAACCTATGACTTAAATTTTGAATAATGAACCTTTTACTATAGAT 1800
    |||||
Db 1741 ATGAAGATTAACCTATGACTTAAATTTTGAATAATGAACCTTTTACTATAGAT 1800
QY 1801 AATTAACGCTATGTTTGTGTCATTAATGACACGCTTACACTGTGATAGTCAATTTT 1860
    |||||
Db 1801 AATTAACGCTATGTTTGTGTCATTAATGACACGCTTACACTGTGATAGTCAATTTT 1860
QY 1861 TCTGCAAAATTAATTAATGAAATTCGCTACTATCAATAGAAAGACGCTGAGATTT 1920
    |||||
Db 1861 TCTGCAAAATTAATTAATGAAATTCGCTACTATCAATAGAAAGACGCTGAGATTT 1920
QY 1921 ACATTTTAATTTAAGCAAAATTTTGAATAATGTTATATTTCTTACATATTTATTA 1980
    |||||
Db 1921 ACATTTTAATTTAAGCAAAATTTTGAATAATGTTATATTTCTTACATATTTATTA 1980
QY 1981 AATATGATGCTATTAATGATTTCTATGTTCTTAAATATTTTAAATTTTATTTAGT 2040
    |||||
Db 1981 AATATGATGCTATTAATGATTTCTATGTTCTTAAATATTTTAAATTTTATTTAGT 2040
QY 2041 TAATATCATTAATGACCAATTAATAGTTGTCATTAATCTCCCTTATATTTTGG 2100
    |||||
Db 2041 TAATATCATTAATGACCAATTAATAGTTGTCATTAATCTCCCTTATATTTTGG 2100
QY 2101 AATATCTACAAATTAATTAATTTAGTCAATTAATCATAGAAAGTTCCAAAAAATTT 2160
    |||||
Db 2101 AATATCTACAAATTAATTAATTTAGTCAATTAATCATAGAAAGTTCCAAAAAATTT 2160
QY 2161 TTGTTAAACAGAACTTCCAAATTTTAAATGAGAACAGAAATTAACAGATAGAAA 2220
    |||||
Db 2161 TTGTTAAACAGAACTTCCAAATTTTAAATGAGAACAGAAATTAACAGATAGAAA 2220
QY 2221 CTATTTTGTGTTGGAATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2280
    |||||
Db 2221 CTATTTTGTGTTGGAATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2280
QY 2281 AGCCTATACGCGCTCAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2340
    |||||
Db 2281 AGCCTATACGCGCTCAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2340

```

```

QY 2341 TCAGAAATGGGACACATGATAAAACGGATTAATTAATTAATTAATTAATTAATTAAT 2400
    |||||
Db 2341 TCAGAAATGGGACACATGATAAAACGGATTAATTAATTAATTAATTAATTAATTAATTA 2400
QY 2401 TTGAGTAAATGTGTTTCTGACTATTAAGGGGCAAAAAAGCAATGCCAAAGTCTAC 2460
    |||||
Db 2401 TTGAGTAAATGTGTTTCTGACTATTAAGGGGCAAAAAAGCAATGCCAAAGTCTAC 2460
QY 2461 GGGTTGACTGTCTCAGTTCCGTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
    |||||
Db 2461 GGGTTGACTGTCTCAGTTCCGTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
QY 2521 TAGGGGTCTCTGACATTTTCACTGTCTTACCCCTACTGCTGAGCCACCTTTTCCA 2580
    |||||
Db 2521 TAGGGGTCTCTGACATTTTCACTGTCTTACCCCTACTGCTGAGCCACCTTTTCCA 2580
QY 2581 TATCTTAAGGTAATTTTGGAAATCCCAATTTTAAACGATGAGACCGTACCGACTTCC 2640
    |||||
Db 2581 TATCTTAAGGTAATTTTGGAAATCCCAATTTTAAACGATGAGACCGTACCGACTTCC 2640
QY 2641 TGGGATTCGCTGAGCATTTATCAAAAATTAATTAAGACGAATGGGTTTATTAATTA 2700
    |||||
Db 2641 TGGGATTCGCTGAGCATTTATCAAAAATTAATTAAGACGAATGGGTTTATTAATTA 2700
QY 2701 AACTCACAACTTGATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2760
    |||||
Db 2701 AACTCACAACTTGATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2760
QY 2761 ATCTAATGACTTTTCTTCTACACAGGTGATGAAAGTTATAGTCTTTAGCCAGAG 2820
    |||||
Db 2761 ATCTAATGACTTTTCTTCTACACAGGTGATGAAAGTTATAGTCTTTAGCCAGAG 2820
QY 2821 ACAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
    |||||
Db 2821 ACAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
QY 2881 CTATTTTCAGATCGCAGCTTCTGCAACTTTTGTGTTTAAATTAATTAATTAATTA 2940
    |||||
Db 2881 CTATTTTCAGATCGCAGCTTCTGCAACTTTTGTGTTTAAATTAATTAATTAATTA 2940
QY 2941 AAGTATTAAGAGGACGCTTAACGAGGCAACAAAGTATTAAGAACGAGAAACCA 3000
    |||||
Db 2941 AAGTATTAAGAGGACGCTTAACGAGGCAACAAAGTATTAAGAACGAGAAACCA 3000
QY 3001 TGAAGCTCATTTGTTAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3060
    |||||
Db 3001 TGAAGCTCATTTGTTAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3060
QY 3061 ATAACTATTAATTTCTGACTTCTTAAACCCCTTACAAACAGAAAGCTCCCTTTT 3120
    |||||
Db 3061 ATAACTATTAATTTCTGACTTCTTAAACCCCTTACAAACAGAAAGCTCCCTTTT 3120
QY 3121 TCACTTAAGATCGGATTCGCAATTTTAAGACAAAGGCAATTAAGAAAGAGTGA 3180
    |||||
Db 3121 TCACTTAAGATCGGATTCGCAATTTTAAGACAAAGGCAATTAAGAAAGAGTGA 3180
QY 3181 GAGAGAGAGAAACTAGCTCC 3202
    |||||
Db 3181 GAGAGAGAGAAACTAGCTCC 3202

```

RESULT 2
AAL15210/c
ID AAL15210 standard; cdna; 883 BP.
XX
XX AAL15210;
XX
XX 07-DEC-2001 (first entry)
XX
XX Human breast cancer expressed polynucleotide 7667.
XX
XX Human; breast cancer; cell marker; cytosstatic; ss.
XX

CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX
 SQ Sequence 9810 BP; 2961 A; 79 C; 2101 G; 4669 T; 0 other;

Query Match 2.6%; Score 84.6; DB 24; Length 9810;
 Best Local Similarity 44.0%; Pred. No. 0.00015;
 Matches 407; Conservative 0; Mismatches 514; Indels 4; Gaps 1;

QY 1527 TATTAAATGCTTCTTATATGACCTTAACTATTATTTAGTTGGAATTAATA 1586
 DB TTTAAAGGGAATGATGGAATATATGATTAATAAATAATATATATATATAT 7905
 QY 1587 GAGCGAAGCTTGACATTTACATATTTATATGATAGTATGATTAATCCAAATA 1646
 DB TATTTTATGAGAGTAAATTTTATTTTATTAAGTTATATTAAGTAATGAAT 7965
 QY 1647 CATACTTGTGATTTAACTTAATCTGTTCTCCFACGGTAATAATTAATCATCG 1706
 DB TATTTTATGAGAAATAGAGAAATGATGAATAAGAAATTTGTTATTTTATGAA 8025
 QY 1707 AGGTAAAAAAGTTTGTCTTATTTTCGGATCATGAAGATTAACCTAATGACTTAA 1766
 DB ATTTATTTGATGTTTATGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 8085
 QY 1767 TTTTGGAAATGTAACCTTTTACTCATGATTAATTAACCGATGTTTGTGCCATA 1826
 DB TATTTTATTAATAAATTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 8145
 QY 1827 ATGACAGCCTCTACACATGATGATCAATTTTCTGCAAAATTAATTAAGAAATCA 1886
 DB TCGAATATATATATATATATATATAGATTAAGATTTTGTGTTTACGAAAAATGTAATTT 8205
 QY 1887 ATGCTACTATCAATAGAGAAACAGCTAGTATTAATTTTAAGACAAATTTT 1946
 DB AATTGAGAGATGATGATTAATAAATAATATATTTTATATATTTTATTAATATA 8265
 QY 1947 TCGAAATGTTATATTTTCTACAAATATTAATTAATGATGATTTTCTT 2006
 DB TATATATGTTATATATATATATATTAATAAGATTAATTTTATTAATTTTATTA 8325
 QY 2007 ATGCTTAAATATTTTATTTTATTTAGTATTAATTAATTAATTAATTAATTA 2066
 DB TATATGTTATATATATATATATGTTATATATATATATATATATATATATATAT 8385
 QY 2067 TGGTGAATTCAAATATCTCATTAATATTTTGAATCTACAAATTAATTAATTA 2126
 DB TGTAT 8445
 QY 2127 CATATACATGCTATGAAGTTCCAAAAAATTTTGTATACGAAACTTCCAAATTTT 2186
 DB AT 8505
 QY 2187 TTTTATATG---GAACAAGATTAACAGATGAAGAACTATTTGTTGATGAGAG 2242
 DB AT 8565
 QY 2243 TAT 2302
 DB TAT 8565
 QY 2303 GTATATCTAGAGTAT 2362
 DB TAT 8685
 QY 2363 AACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2422
 DB AAAAAGATGATATTTTATTTTATATATATATATATATATATATATATATATATATAT 8745

QY 2423 TATTTGAGGGGCAAAAAAGACAAAT 2447
 DB 8746 GTAT 8770

RESULT 4

ABK28257/c
 ID ABK28257 standard; DNA; 8170 BP.

AC ABK28257;

XX 23-APR-2002 (first entry)

DE DNA transcription associated genomic DNA #66.

XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angiodysplastic congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamine disorder; solid tumour.

XX Unidentified.

XX WO200192565-A2.

PD 06-DEC-2001.

XX 06-APR-2001; 2001WO-EP03973.

XX 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-090046/12.

DR New nucleic acids or oligomers, useful for diagnosing or treating

PT diseases associated with DNA transcription, e.g. immunological

PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid

PT tumours or cancer

XX Claim 1; SEQ ID No 131; 32pp; English.

XX The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABK28117-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention.
 CC Note: The sequence data for this patent did not form part of the printed

OY 1937 ACAAAATTTTGGAAATGTTATTAATTCACATATTTAAATATGATGCTATA 1996
 DB 488 TATTTTGAATTAATTAATGATGTTATTTTGTGAAATTTTGTGAAAT 547
 OY 1997 TGTATTTCCATATGTTCTTAATAATTTTATTTTATTTAGTTATTAATATGATGAC 2056
 DB 548 TTTATTTATTAATGATTTATTTGTTATTTATTTAGTTATTAATTTTATTTTATTTTAA 607
 OY 2057 CAATATAGTTGCGAATTCACAAATGTCATTAATTTTGTGAAATTCACAAATAT 2116
 DB 608 TATGATTTTATTTTATTAATTAATTTAT -ATTTTATTTTATTTTATTTTATTTTATTTA 665
 OY 2117 AATATTTAGTCAATACAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2176
 DB 666 AATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 723
 OY 2177 CCAATTTTATTTTATTTATGAAACAAAGAAATGATGATGATGATGATGATGATGATG 2236
 DB 724 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 783
 OY 2237 TGGAAATGATTAATATACATTAACCAATTTTAAATTTATTAATGATGATGATGATGAT 2296
 DB 784 ATGATATTTATTAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTT 838
 OY 2297 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 898
 DB 839 AATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 958
 OY 2357 AATGAAAGGAAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2416
 DB 899 GAAAAAGGAAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 958
 OY 2417 TCTGACTATTTGAGGCAAAAAAAGCAATGCCAAA 2454
 DB 959 GATTAAGAAAAATAGTAAATAATATATTTTGAAGA 996
 RESULT 6
 ID ABL32426
 AB32426 standard; DNA; 9810 BP.
 AC ABL32426;
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 399.
 XX Human immune system associated gene
 KW Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antiasthmatic; anti-HIV; anticonvulsant; ophthalmological;
 KW antileukemic; antidiabetic; antidiabetic; antidiabetic;
 KW antileukemic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 OS Homo sapiens.
 PN WO200200928-A2.
 XX 03-JAN-2002.
 PF 02-JUL-2001; 2001WO-EP07537.
 XX 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIC-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI: 2002-130909/17.
 DR

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PS Claim 1: SEQ ID NO 399; 32pp + Sequence Listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 SQ Sequence 9810 BP; 2770 A; 79 C; 1978 G; 4983 T; 0 other;
 Query Match 2.68; Score 82.6; DB 24; Length 9810;
 Best Local Similarity 46.98; Pred. No. 0.00031;
 Matches 374; Conservative 0; Mismatches 404; Indels 19; Gaps 3;
 OY 1625 TAGATGATGATTTCCAAATACATCTGATGATGATGATGATGATGATGATGATGATG 1684
 DB 822 TTTGTTGAATGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 881
 OY 1685 ACGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1741
 DB 882 GTGAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 941
 OY 1742 TGAAGATTAACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 941
 DB 942 TTTTATTTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1001
 OY 1802 ATTACGATGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 1061
 DB 1002 AATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1061
 OY 1862 CTGCAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1921
 DB 1062 TATATATTAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 1121
 OY 1922 CATTTATTTAAGCAAAATTTTGAATA-----ANGTATTAATTTCT 1966
 DB 1122 TTTTATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAT 1181
 OY 1967 AACATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2026
 DB 1182 ATAAATTTATGATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1241
 OY 2027 TTTATATTTAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2086
 DB 1242 ATATATTAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2101
 OY 2087 ATTAATTTTGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2146
 DB 1302 ATATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAT 1361
 OY 2147 TTTCAAAAAATTTTGTGAAGAACTTCAAAATTTTGTGATGATGATGATGATGATGATGAT 2205
 DB 1362 ATATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAT 1421
 OY 2206 ATTAAGATTAAGAAATTAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2265
 DB 1422 AATATATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAT 1481
 OY 2266 TTAATAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAT 2325
 DB 1482 ATGATATTAATAATTTGTTAAAGGATTAATTAATTTATTAATTAATTAATTAATTTAT 1541
 OY 2326 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2385
 DB 1542 GTATATTAATTAATAATTAATTAATTAATTAATTTATTAATTTATTTATTTATTTATTT 1601

QY	2386	AAATATATATATATATTT	2402
DB	1602	AATTAGCATGTATTTT	1618
RESULT 7			
AA545388/c			
ID	AA545388	standard; DNA; 15732 BP.	
XX			
AC	AA545388;		
XX			
DT	18-DEC-2001	(first entry)	
XX			
DE	Chemically pretreated genomic DNA associated with cell cycle #47.		
XX			
KM	Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;		
KM	human immunodeficiency virus; neurodegenerative disorder; solid tumour;		
KM	grat-versus-host disease; glomerular disease; lewy body disease; cancer;		
KM	arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;		
KM	immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;		
KV	PCR primer.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200168911-A2.		
XX			
PD	20-SEP-2001.		
XX			
PF	15-MAR-2001; 2001WO-EP02945.		
XX			
PR	15-MAR-2000; 2000DE-1013847.		
PR	06-APR-2000; 2000DE-1019058.		
PR	07-APR-2000; 2000DE-1019173.		
PR	30-JUN-2000; 2000DE-1033529.		
PR	01-SEP-2000; 2000DE-1043826.		
XX			
PA	(EPiG-) EPIGENOMICS AG.		
XX			
PI	Olek A, Piepenbrock C, Berlin K;		
XX			
WPI:	2001-602751/68.		
XX			
PT	Designing primers and probes for analysing diseases associated with		
PT	cytosine methylation state e.g. arthritis, cancer, aging,		
PT	arteriosclerosis comprising fragments of chemically modified genes		
PT	associated with cell cycle -		
XX			
PS	Clalm 1; SEQ ID No 93; 28pp; English.		
XX			
CC	Sequences AA545296-AA545520 represent chemically pretreated genomic DNA		
CC	molecules associated with the cell cycle and specific PCR primers of the		
CC	invention. The sequences are useful for detecting the methylation state		
CC	of all CpG dinucleotides in a sequence and therefore for analysing		
CC	associated diseases. By analysing cytosine methylations in the pretreated		
CC	DNA, genetic and/or epigenetic parameters for the diagnosis and therapy		
CC	of existing diseases or the predisposition to specific diseases can be		
CC	ascertained. The parameters may be compared to another set of genetic		
CC	and/or epigenetic parameters, the differences serving as basis for		
CC	diagnosis and/or prognosis events which are disadvantageous to patients.		
CC	The sequences of the invention are useful for the diagnosis and therapy		
CC	of HIV infection, neurodegenerative disorders, graft-versus-host disease,		
CC	aging, glomerular disease, lewy body disease, arthritis,		
CC	arteriosclerosis, solid tumours and cancers.		
XX			
SO	Sequence 15732 BP; 4638 A; 70 C; 2672 G; 8352 T; 0 other;		
Query Match	2.6%;	Score 82.6;	DB 22; Length 15732;
Best Local Similarity	43.9%;	Pred. No. 0.00032;	
Matches	453; Conservative	0; Mismatches	574; Indels 6; Gaps 2
QY	1400	AAACAATATAAACAAACAGAAAAAGAAACGATTTTCTTGCGATTTCATTCAATGA	1459

[illegible]

PS Claim 1; SEQ ID No 148; 27pp; English.

XX The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes. Sequences with even numbered Seq ID numbers are the
CC complementary sequence of the corresponding odd numbered sequence
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5689 BP; 1490 A; 207 C; 1277 G; 2707 T; 8 other;

Query Match 2.5%; Score 81.6; DB 22; Length 5689;
Best Local Similarity 49.1%; Pred. No. 0.00044;
Matches 285; Conservative 0; Mismatches 292; Indels 3; Gaps 3;

QY 1824 ATATGACAGCCTCTACACGTGATGCAATTTTTCGCAATATTAATTAAGCAAT 1883
DB 2110 AAAATGCCAATCCGCAACCGCAATATCAAAAAATTTCCGCCGAAAAA 2051
QY 1884 TCAATGCTACTATCAATGAAAGACGAGATATACATTTTAATTAAGCAAAAT 1943
DB 2050 TATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1991
QY 1944 TTTTGAATAATGTTAATTTCTACAAATTTTAATTAATGCTAATATTTT 2003
DB 1990 ATTATTTAAAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAA 1931
QY 2004 CCGATGTTCTAAATTTTCTTATATTTGTTAATTAATCAATATGACCAATAT 2063
DB 1930 NATAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1871
QY 2064 AGTTGGT-GAATTCAAATATCTCCATTAATTTTGAATGCTACAAATTAATAT 2122
DB 1870 AATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1811
QY 2123 TACTCAATTAACATGCAT-AGAAAGTCCAAAAAATTTGTTAACAAGAACTTCCAA 2181
DB 1810 TATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1751
QY 2182 TTTTGTTTTGTGGAACAGAAATTAACAGATAGAAACATATTTGTTGCGATGAA 2241
DB 1750 TAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1691
QY 2242 GTATGTAATTAACATTAAGCAATTTTAAATTAATAGCTATAGCGCTCAAGTA 2301
DB 1690 TAACAATAAATAAATAGCAATATTAATTAATTAATTAATTAATTAATTAAT 1631
QY 2302 TGTATCTAGTGTGCTATTAATA-ATGCATAGTGCGGTTAGAAATGGGCAACAATG 2360
DB 1630 TATTAATCTTCAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1571
QY 2361 AAAACGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2400

DB 1570 AATAAAAAATTTTAATTAACGTTAAAAAATATTAATATAT 1531

RESULT 11
ID ABR28226/c
ID ABR28226 standard; DNA; 5689 BP.
XX ABR28226;
XX 23-APR-2002 (first entry)
XX
XX DNA transcription associated complementary genomic DNA #50.
XX
XX PNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
XX DNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
XX single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
XX viral infection; Sezary syndrome; haematological disorder; tuberculosis;
XX immunological disorder; Werner syndrome; developmental disorder;
XX psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
XX neurodegenerative disorder; Maardenburg syndrome; Niemann-Pick disease;
XX myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
XX angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;
XX polyglutamine disorder; solid tumour.
XX Unidentified.
XX OS
XX PN WO200192565-A2.
XX PD 06-DEC-2001.
XX PE 06-APR-2001; 2001WO-EP03973.
XX PR 06-APR-2000; 2000DE-1019058.
XX PR 07-APR-2000; 2000DE-1019173.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPIC-) EPIGENOMICS AG.
XX PI Olek A, Plepenbrock C, Berlin K;
XX DR WPI; 2002-090046/12.
XX
XX New nucleic acids or oligomers, useful for diagnosing or treating
XX diseases associated with DNA transcription, e.g. immunological
XX disorders, Werner syndrome, psoriasis, myocardial infarction, solid
XX tumours or cancer
XX
PS Claim 1; SEQ ID No 100; 32pp; English.

XX The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Maardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
CC or cancer. Sequences ABR28127-ABR28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.

ABA92787
ID ABA92787 standard; DNA: 640681 BP.
XX
XX ABA92787;
AC
XX 27-MAR-2002 (first entry)
DT
XX Buchnera sp. genomic DNA SEQ ID NO:1.
DE
XX Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
KW circular; ds.
XX
XX Buchnera sp.
OS
PN JP2001292771-A.
PD 23-OCT-2001.
PE 07-APR-2000; 2000JP-0107160.
PR 07-APR-2000; 2000JP-0107160.
XX
XX (RIKA) RIKAGAKU KENKYUSHO.
PA
DR WPI: 2002-126043/17.
XX
XX A genomic DNA of cockroach-symbiotic bacterium -
PS Claim 1; Page 16-230; 237pp; Japanese.
XX
XX The present invention describes a gene (I) derived from Buchnera sp.
CC containing the DNA (a) or (b), (a) has a fully defined base pair
CC sequence selected from a table of sequences found in the Buchnera sp.
CC genomic DNA of ABA92787 given in the specification or is a DNA selected
CC from complementary DNA sequences, and (b) is a DNA which hybridises with
CC the DNA (a) and encodes a protein. Also described are: (1) a recombinant
CC vector (II) containing (I); (2) a transformant (III) containing (II);
CC (3) a genomic DNA of Buchnera sp. containing the sequence given in
CC ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
CC (d), (c) is a DNA containing a fully defined sequence given in ABA92788
CC or ABA92789 and (d) is a plasmid which hybridises with a DNA; and (5) a
CC method for the preparation of a protein in which (III) is cultured and
CC the expression protein of the objective protein is collected from the
CC resultant culture. The DNA is useful for developing agricultural
CC chemicals for exterminating cockroaches. The present sequence represents
CC the specifically claimed Buchnera sp. genomic DNA sequence, from the
CC present invention.
XX
XX Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;
SO

Query Match 2.5%; Score 80; DB 24; Length 640681;
Best Local Similarity 45.8%; Pred. NO. 0.00086;
Matches 395; Conservative 0; Mismatches 455; Indels 12; Gaps 3;
Query 1559 ATATTATTGTTAGTGAATTTATAGAGGAGCTGTACATTACATTTATATT 1618
DB 324686 ATATTGTTAGTATCTTTTATATTACCAATATTATTCATTAATTTGGCAAC 324745
QY 1619 AGATACAGTATGCTGATTTCCAAATACACTTTGGATGTTAACTTATCTGTTT 1678
DB 324746 AATCTTTTGTACAAAGAAATATATTTTATTTATCATATAAATCCATTTTGATC 324805
QY 1679 CTTCCTACGATTAATATTAATCATCGAGTAAAGAAAGTTTGTCTTATTTTGGCAT 1738
DB 324806 TCTTCTATCATGTAAATAATCAATTTCTTCTTAATATCAATTAATCTTTTATATAT 324865
QY 1739 GCATGAAGGATAACCTATGACTTATTTTTCGAATGTAACCTTTTACTATAGA 1798
DB 324866 TGTGGCTGATTTTATGCTCATTAATAATCTTAAATTTTAAAGATTTTCTTCT 324925
QY 1799 TTAATACGATGTTTGTGCAATATGACACCTCTACAACTGATAGTCAATTT 1858
DB 324926 GTCAAGAAATATTTATTTTTCGTAAGAAATGATATAATGACAAATACGATATTT 324985

QY 1859 TTTCGCAAAATATTAATTAGCAATTCAGTCTACTATCAATAGAGAAACAGCTGAGTA 1918
DB 324986 TTTTTCGATTTATTTACATTAATAAATACATCAGTAATTTTGANATTTAAATTAATA 325045
QY 1919 TTACATTTTAAATTTAAAGCAAAATTTTGGAAAAATGTTATTTTCTACATATTTAT 1978
DB 325046 ATTTTGTGTTTAAATTTTAAATTTTGTAAATTTTCATGAAACGTATTTCTTAAAT 325105
QY 1979 AAAATATGATGCGCTTAATGATTTCCATA- - - - -TCTCTTAAATATTTTATATAT 2033
DB 325106 AAGATTAATTTTATTTTAAATTTTAAATTTTAAATGCTATTAATTTTATATTA 325165
QY 2034 TTAGTTAATAATATATATTAATGAAACCAATTAATAGTGGTGAATCAATATCTCATTAATA 2093
DB 325166 ATTTAAATATTTCTTAAATTAATTTTAAATGCTATTAATTTAAATTAATAATA 325225
QY 2094 TTTTGTGAATCTACAAAT- - - - -TATTAATTTTATGTAATTAACAAATGCAATGAAAGTT 2148
DB 325226 ATTTATTTTATTTTACATTTTATTAATTAATTAATTAATTAATTAATTTTATTA 325285
QY 2149 CCAAAAAATTTTGTGTTAAGCAAACTTCCAAATTTTATTTTATGGAACAGAAATA 2208
DB 325286 ATTAATAATTCATTTTATATTAATAATTAATTAATTAATTTTATTTTATTTTAAAT 325345
QY 2209 ACAGATGAGAAATCTATTTTGTGGAATGAGATTAATTAATTAATTAATTAATTTTAA 2268
DB 325346 ACTATTAATGATA- - - - -ATTATAGATATTAATTAATTAATTAATTAATTAATTAATTA 325403
QY 2269 AAAATTTATTAAGCTATACGCGCTCAAAAGTATGATCTAGTGGTATTAATTAAT 2328
DB 325404 TCAATTAATTAATTAATTAATTAATTAATTTCTTTTATTTATTAATTTTATTAATA 325463
QY 2329 GCATGTCGATTCAGATTTGGGACACAAATGAAACGAATTAATAATTAATTAATTAATA 2388
DB 325464 GAATTTGCATGCAATCGAATTAATAATTAATTTTATTTTCAATTAATTAATTAATA 325523
QY 2389 ATTAATTAATAATTTGATTAAT 2410
DB 325524 TTTCAATTAATAATTTTATTAATTTTATTAATTTTATTAATTTTATTAATA 325545

RESULT 15
ABK31511/c
ID ABK31511 standard; DNA: 47108 BP.
XX
XX ABK31511;
AC
XX
XX 23-APR-2002 (first entry)
DT
XX
XX Signal transduction associated gene modified complementary DNA #177.
DE
XX Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
XX Homo sapiens.
OS
XX Synthetic.
PN WO200200926-A2.
PD 03-JAN-2002.
PE 29-JUN-2001; 2001WO-EP07472.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Phippenbrock C, Berlin K.
XX
XX WPI: 2002-147896/19.

XX Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
XX sequences of genes associated with signal transduction
PS Claim 1: SEQ ID No 354; 24bp; English.

CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (CpG islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK1158-ABK11545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
XX European Patent Office.

XX Sequence 47108 BP; 15349 A; 429 C; 8736 G; 22594 T; 0 other;

Query Match

Best Local Similarity 2.5%; Score 79.6; DB 24; Length 47108;
Matches 457; Conservative 0; Mismatches 564; Indels 9; Gaps 3;

QY 1074 TTCACGCTTTATTTTACAAAGTACAGATATAGCTTGAATCCCTTGAATAATTTT 1133
DB 1033 TTAATTCCTTTACCAAAAAACATTTATATATTAATTAACCAAACTTAATAATCT 974
QY 1134 GGAATTCGCGTTGTTATGTAATATAGTTAGTGTAAACAATATGTAATCAATTA 1193
DB 973 AAACCTTAAAAATTAATTAATATCAAAATTCATTTCTTAATTTAATTAATTAATA 914
QY 1194 GTGGTCAACAT 1253
DB 913 ATTAAT 854
QY 1254 CATATATGCGTAT 1310
DB 853 TATTCCTCAAT 794
QY 1311 ATGTGATGCGTAT 1370
DB 793 AAATTAAT 735
QY 1371 ATCACTCAAT 1430
DB 734 TTTCTTAAAT 675
QY 1431 AACGATTTTCTGATTCATTCATGATTAATATATATATATATATATATATATATAT 1490
DB 674 AAATCCAT 615
QY 1491 TTGGAAGTCTCTACAGGCTGTAACCATCTGGAATTAATATATATATATATATAT 1550
DB 614 TCCCAAT 555
QY 1551 TCTTAACAT 1610
DB 554 AT 495
QY 1611 TTTAT 1665
DB 494 TAT 435

QY 1666 CTATATCTGTTTCTCTACAGGATTAATATATATATATATATATATATATATATAT 1725
DB 434 TTACCTTTAACCTTTAAAAATTTTACTTATTTTATTTTCTCTCAATATCTAATTT 375
QY 1726 TTTATTTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1785
DB 374 TCAATCTCAAT 315
QY 1786 TTTTCTCAAT 1845
DB 314 TTTTCTCAAT 255
QY 1846 TGAATGCTAATTTTCTGCAAAATATATATATATATATATATATATATATATAT 1905
DB 254 TTAAT 195
QY 1906 AAACGCTGAT 1965
DB 194 ATTAATCTCAAT 135
QY 1966 TTAACAT 2025
DB 134 CTAAAT 75
QY 2026 TTTTAT 2085
DB 74 CTAAAT 15
QY 2086 CATTAATAT 2095
DB 14 TTTTAATTT 5

Search completed: March 29, 2003, 18:59:11
Job time: 2214.58 secs

Handwritten scribbles and marks, possibly representing a signature or initials.

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:09:40 ; Search time 111.977 Seconds
(without alignments)
8769.450 Million cell updates/sec

Title: US-09-502-426a-1_COPY_1_3202
Perfect score: 3202
Sequence: 1 atgtgggtatattatgttgc.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77.4	2.4	19124	2	US-08-487-826B-13
2	68.6	2.1	837	4	US-08-998-416-288
3	67.4	2.1	19124	2	US-08-487-826B-13
4	65.8	2.1	615	4	US-08-998-416-186
5	65.4	2.0	636	4	US-08-998-416-1137
6	65.4	2.0	20674	4	US-09-641-638-651
7	63.6	2.0	8920	2	US-08-446-855A-1
8	63.6	2.0	8920	4	US-09-150-741-1
9	63.6	2.0	20674	4	US-09-641-638-651
10	62.6	2.0	4818	4	US-08-817-926-27
11	61.8	1.9	5852	1	US-07-867-106-2
12	61.2	1.9	1431	4	US-09-316-083-2
13	60.8	1.9	4818	3	US-08-817-926-27
14	60.6	1.9	837	4	US-08-998-416-288
15	60.4	1.9	8920	4	US-08-446-855A-1
16	60.4	1.9	8920	2	US-09-150-741-1
17	60.2	1.9	665	2	US-08-883-795A-36
18	60	1.9	1864	4	US-09-468-265-4
19	59.8	1.9	6124	4	US-08-213-419B-3
20	59.2	1.8	51952	3	US-08-947-823-1
21	59	1.8	827	4	US-08-998-416-535
22	58.4	1.8	665	2	US-08-883-795A-36
23	58.2	1.8	12793	4	US-09-004-838-124
24	58.2	1.8	6124	4	US-08-213-419B-3
25	57.8	1.8	1511	1	US-07-991-867B-8
26	57.8	1.8	1511	1	US-08-107-755A-8
27	57.8	1.8	1511	2	US-08-544-332-8

c	28	57.8	1.8	1511	4	US-09-370-861A-8	Sequence 8, Appl1
c	29	57.4	1.8	3095	4	5231168-1	Patent No. 5231168
c	30	56.6	1.8	168575	6	US-09-426-290-1	Sequence 1, Appl1
c	31	56.4	1.8	2058	4	US-08-749-391-1	Sequence 1, Appl1
c	32	56.4	1.8	2058	3	US-09-390-200-1	Sequence 1, Appl1
c	33	56.4	1.8	4673	5	US-07-638-431-1	Sequence 1, Appl1
c	34	56.4	1.8	4673	5	PCT-US92-00018-1	Sequence 1, Appl1
c	35	56.2	1.8	1431	4	US-09-316-083-2	Sequence 2, Appl1
c	36	56.2	1.8	2960	3	US-08-913-842-3	Sequence 3, Appl1
c	37	55.6	1.7	701	4	US-08-998-416-701	Sequence 701, App
c	38	55.6	1.7	5852	1	US-07-867-106-2	Sequence 2, Appl1
c	39	55.2	1.7	1406	3	US-08-913-842-6	Sequence 6, Appl1
c	40	55.2	1.7	1511	1	US-07-991-867B-8	Sequence 8, Appl1
c	41	55.2	1.7	1511	1	US-08-107-755A-8	Sequence 8, Appl1
c	42	55.2	1.7	1511	2	US-08-544-332-8	Sequence 8, Appl1
c	43	55.2	1.7	1511	4	US-09-370-861A-8	Sequence 8, Appl1
c	44	55	1.7	1850	3	US-08-617-860B-32	Sequence 32, Appl1
c	45	55	1.7	4098	2	US-08-605-106-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-487-826B-13

Sequence 13, Application US/08487826B
Patent No. 593827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhuan

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,826B

FILING DATE: 10-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Isreelsen, Ned

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121,001CPI

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 19124 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: linear

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-487-826B-13

Query Match 2.4%; Score 77.4; DB 2;
Best Local Similarity 42.6%; Pred. No. 1.8e-06; Length 19124;

Matches	590;	Conservative	0;	Mismatches	781;	Indels	14;	Gaps
QY	1013	TTAAAAACGATTTTACACCGCTACCTAGTGTGAGATTACTAGCATTAAGCATTAAGGACCC	1072					
Db	229	TTATATATCAAAATTTATATATAGTTCCTATTAATAAATAATATATATATACAAATTAAT	288					
QY	1073	GTTCAAGCATTTTATACAAAGTTACAAACTGAAATAGCTGAAATCCCTTGAAAAATTT	1132					
Db	289	TTCTTGTTATTTTATATTAATATATACCTAAATTCCTATTTTATTTTATTTATCTTAATTCCTTTT	348					
QY	1133	TGGAATTTACCGGTTGTATATTAATATATAGATTAGTGTAAACAATATATGTTAATCAAT	1192					
Db	349	AATTTCTTAATTCCTTTATATGCAACAAAAAACAATAAAGTAATTCATCATATCAACAAAA	408					
QY	1193	ACTGGTCAACATATATCATATATTCCTTACGAAAAAACAACCTTAAGAGAATTACATAT	1252					
Db	409	AAAAAAAAAAAAAAAAAAAAAAAAATTTATATATATTAATTAATTAATTAATTAATTAATTAAT	468					
QY	1253	CCATATATGGGTATGCTATACCTTCACGTAATGCTATACTACAGACTAAGAAATGTTAT	1312					
Db	469	TCACCTTATTTATTTAATATGATTTTATACATTTAAACATATTTGAGATTTTAATTAATTA	528					
QY	1313	GTGATGTGATTAATGAATAATTCACACCGCTGTAAATTAATTAATGAGACCTATATGTCAGAT	1372					
Db	529	TTTAAACATAGAAGAGGTTAAGAAATACATTTTTTTTTTTTTTTTGTATGTATATTCACAAAT	588					
QY	1373	CACGTCAAAATATCATCTCTGTGTGTGTCAACAATTAATAAACAACAAAAAGAAAAAGAAAA	1432					
Db	589	AAATATATATATATATATCTTTTAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAAT	648					
QY	1433	CGATTTTCTTGAGATTCATTCATAGATGCTAATTAATGACATAGATCTTTTGCTTACAGTTT	1492					
Db	649	TCTTTTATC-----ACATGTGAAAAATTTAAATAATTTTCGATTTTATTCATATATTTT	702					
QY	1493	CGAAGTCCCTTCACAACCGGTGACCATCGACATCTTAATTAATGCTTTCTTTAATATGCATC	1552					
Db	703	ATGCTGTTATATATCTTATATATAGCTTTATATACATATGATTAATTAAGAAGTAAATAGCCT	762					
QY	1553	TTTAAACATATTTATTTGTTAGTTGGAATTTAATTAAGAGCAACTGTACATTACAATAT	1612					
Db	763	AATATATTAATAATACGATCGTATTTTAAATTTCACTTATATATTTCAAAATATATTTCCATGT	822					
QY	1613	TATATTAATTAATTAATGATGATTTATTCCAATTAATTAATTAATTAATTAATTAATTAATTAAT	1672					
Db	823	TTATTTTCAAAATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	882					
QY	1673	TTGTTTCTTCACACGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT	1732					
Db	883	GCATTTACATGACATTTTAAATTTTATTAATAAAAAATTTTTTTTAAATATTAATTAACA	942					
QY	1733	CGCGATGACGAGGATTAACCTTAATGACTTAATTTTGTGAATAATGTAACCCCTTTTACT	1792					
Db	943	AATTTCAATACATTAATTAATTTTACACACAACATTTAAGTTGTCATTAATGTAACATTAAT	1002					
QY	1793	CATAGATTAATTAACCGTATGTTTTTTGTGCGCAATTAAGACAGCCTCACACCTGTATAGT	1852					
Db	1003	AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1062					
QY	1853	CAATTTTTCGCAAAATTTAAATTAAGGAATTCATAGCTATCAATGTAATGAAGAAACAGC	1912					
Db	1063	CTGTATTTCAATAGTATATATTAATTTGTTATTAACGCTTCAAAATATATATTAATTAATTA	1122					
QY	1913	TGAGTATTAATTAATTAATTAATTAAGCAAAATTTTGAAGAAATGTAATTAATTTCTAACAAT	1972					
Db	1123	TTAAAAATATATATATATATTAATTAATTTTGTATGTATAGTATATATATGCAATTAATA	1182					
QY	1973	ATTATTAATAATATGATGCTATATATATATTTCCATATGTTCTTAAATATTTTTTTTATA	2032					
Db	1183	TAAAGTAAAAATTTCTATACCTATTAATTAATTAATTAATTTATATATATATATATATA	1237					
QY	2033	TTTATGTTAAATACATTAATTAAGCAAAATTAATGTTGGGAATTCGAATATATTCATTAAT	2092					
Db	1238	TATATATATAGTATATGTTATCAAAATTAATTAATTAATTAATTTATGTAATTT-----ATTAAATTAATTA	1294					

```

OY 2093 ATTTTGGAAATCTACAAATATATATATTTAGTCAATTAACATGCATAGAAAGTCCAA 2152
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1295 TTTGTATATACCTACACAGACTTAAGAAACTATACAACTGCTATCTAATAGTATATAT 1354
    || || || || || || || || || || || || || || || || || || || || || ||
OY 2153 AAAAAATTTGTATACAGAACTCCAAATTTTTTTTTTTTATGAGACAGAAATACAG 2212
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1355 ATATATCTTTTTTATATTTAATTTGTCCTCTCTTTTTTTTTTTTTTAAATAAATTAAT 1414
    || || || || || || || || || || || || || || || || || || || || || ||
OY 2213 ATAGAAACATTTTTTGTGTGGAAATGSAAGTAGTAATATACATTAAAGCAATTTAAAA 2272
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1415 TAAATATTTTTTTTTCATTAATTTATAGATTAGTATTAATTAATTAATTAATCTTTTAA 1474
    || || || || || || || || || || || || || || || || || || || || || ||
OY 2273 ATATATATAGCCCTATACGCCCTCAAGATNGTATCTAGTAGTGATTAATTAATGCA 2332
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1475 AAAACTTCAAAAACATTTTGTGCATTAATAATATATATATTAGTAACCACTTGATTAAT 1534
    || || || || || || || || || || || || || || || || || || || || || ||
OY 2333 GGTGCGATTCAGAAATTTGGACAACAATGAAMACGAATTAATAATTAACCTTAAAAATA 2392
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1535 AGAGAGAAACGTAGACATACACAAAAAATTTGAAACAAAAAGAAATTTACAAAAAATTA 1594
    || || || || || || || || || || || || || || || || || || || || || ||
OY 2393 ATAAA 2397
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1595 TAAAA 1599
    || || || || || || || || || || || || || || || || || || || || || ||

RESULT 2
US-08-998-416-288
Sequence 288, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reblischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
TITLE OF INVENTION: AND USUS THERMOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESS: NO. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Melgs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 288:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
;
US-08-998-416-288

```

Query Match	2.1%;	Score 68.6;	DB 4;	Length 837;
Best Local Similarity	46.9%;	Pred. No. 4.6e-05;		
Matches 250; Conservative	0;	Mismatches 279;	Indels 4;	Gaps 1

[illegible]

RESULT 3
 US-08-487-826B-13/C
 : Sequence 13. Application US/08487826B
 Patent No. 5993827
 GENERAL INFORMATION:
 APPLICANT: Slim, Kim L.
 APPLICANT: Chitnais, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 NUMBER OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 City: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

1 APPLICATION NUMBER: US/08/487,826B
2 FILING DATE: 10-SEP-1993
3 CLASSIFICATION: 435
4 ATTORNEY/AGENT INFORMATION:
5 NAME: Israel, Ned
6 REGISTRATION NUMBER: 29,655
7 REFERENCE/DOCKET NUMBER: NH121,001CPT
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: (619) 235-8550
10 TELEFAX: (619) 235-0176
11 INFORMATION FOR SEQ ID NO: 13:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 19124 base pairs
14 TYPE: nucleic acid
15 STRANDEDNESS: single
16 TOPOLOGY: linear
17 MOLECULE TYPE: cDNA
18 HYPOTHEetical: NO
19 ANTI-SENSE: NO
20 US-08-487-826B-13

Query Match	2.18;	Score 67.4;	DB 2;	Length 19124;
Best Local Similarity	42.38;	Pred. No. 0.00016;		
Matches 821; Conservative	0;	Mismatches 1091;	Indels 29;	Gaps 7

[illegible]

Db 116 ATTATTAATCTATTATTAATATATTTATATATTAATTACTTAATTCATCATATTATATAT 175

Qy 1870 ATTAATATTAGCAATTCATCGCTACTATCAACATAGAGAAACAGCTGATATACATTTTAA 1922

Db 176 TATATTAATTAATTAATAATATTTTAATATGAAATCTATTGACTGTCATTAATTTTA 235

Qy 1930 TTTAAAGCAAAATTTTGGAAAAATGTTATTAATTTCTAACAAATATTTAAATATAG 1980

Db 236 ATTACTGTTTAAAAATATTTATAGATATTTATTTCTTTAAATAAATTTTAAATAGAT 295

Qy 1990 CCTAATAGTATTCGATGCTCTAAATAATTTT---TTTATATTTAGTTATTAAT 2045

Db 296 ATCATATTAATTAATTAATTTATTAATTTGTTATTAATTAATTAATTAATTTTATTTTANA 355

Qy 2046 ACATTAATGAAACCAATTAATAGTGGTGCATTCGAATTCATTCATTAATTTTGAATC 2105

Db 356 AAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTANA 415

Qy 2106 TACAATTAATTAATTTAGTCAATTAACAATGCAATGAAAGTCCCAAAAAATTTGGT 2165

Db 416 AATATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 475

Qy 2166 AACGAAACTGCCAAATTTTTTTTT 2192

Db 476 AACTTAATTTCTTAATTAATTTT 502

RESULT 5

```

US-08-998-416-1137
Sequence 1137, Application US/08998416
Patent No. 6239264

GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Rolf
APPLICANT: Redlschning, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
TITLE OF INVENTION: AND USBS THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38, 241
REFERENCE/DOCKET NUMBER: PE/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8687
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1137:
SEQUENCE CHARACTERISTICS:
LENGTH: 636 base pairs
TYPE: nucleic acid

```

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1692RP
;
DS-08-998-416-1137

```

Query Match	2.0%;	Score 65.4;	DB 4;	Length 636;
Best Local Similarity	47.1%;	Pred. No. 0.00018;		
Matches 269;	Conservative 0;	Mismatches 266;	Indels 6;	Gaps 2

QY	1841	AACTGATGTCGAATTTTTCGCAAAATTTAAATGAGAACTCAATGCTACTACAAAT	1900
Db	29	AAAGATTAATTAACCTTTTATTATTAATATTTTAAGTATTAATTAATTTTAACCTATTATTT	88
QY	1901	AGAGAGAACAGCTGACTATTTTACATTTTAAATTTAAAGACAAAATTTTGGAAAAATGTTATA	1960
Db	89	ATCATATTTTAAATTAATTAATTTATTTGATTTATTAATACCTATTAATTAATTTATTAATA	148
QY	1961	ATTTCAACAATTTTATTATA - AATAGATGCCATATATGTAATTCCTAATGTTCTTAAT	2019
Db	149	TTTACTTTAATTCATCTCAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	208
QY	2020	ATTTTATTTTATTTATAGTTATTAATTAATACATTAATGAAACCAATAATAGTGGTAATCCAA	2079
Db	209	ACGATTTATCTATGTCGCAAAATTTTAAATTTAGTTATTTAAATTTTATTAAGATATTAAT	268
QY	2080	TATCTCCATTAATTTTATTTTGAATCTACAAATTTATTAATTTTATGTCATTAACATGCA	2139
Db	269	TTTCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	328
QY	2140	TGAAAGTCCAAAAAATTTTGTTAACAGAACTGCCAAATTTTATTTTATTTGGA	2199
Db	329	ATTAAA-----ATAATTAATTTTATTAATTAATTAATTAATTAATTTTAAATTTGTAA	383
QY	2200	CAGAAGATACAGATAGAGAAACTATTTTGTGTGGAATGGAAGTAGTAATTAACATTAG	2259
Db	384	TTATATATTTTATTTTAAATATCTATTTTATTAATTAATTAATTAATTAATTTATTTATTAAT	443
QY	2260	CAAAATTTTAAAAAATTAATTAAGCCTTAACGCGTCAACAGTATGTAATCTGATAGCTGA	2319
Db	444	CTTTTATTAAGATTAATTAATTAATTAATTAATTTTAACCTTAATTTTATTTTATTAATTT	503
QY	2320	ATTATATATGATGGTGGCATTCAGAAATTTGGACAAACATGAAAACGGAATTAATTAATTT	2379
Db	504	ATATATTTTAAATTAATTAATTAATTCATTTATTTATTTATTTATTTATTAATTAATTAATTT	563
QY	2380	AACTTTAAATTAATTAATTAATTTTGTAAT 2410	
Db	564	TAATTAATTTTATCATTTATTTAATTAAT 594	

RESULT 6

```

US-09-641-638-651
? Sequence 651, Application US/09641638
? Patent No. 6432648
? GENERAL INFORMATION:
? APPLICANT: Blumenfeld, Marta
? APPLICANT: Bougueterec, Lydie
? APPLICANT: Chumakov, Ilya
? APPLICANT: Cohen, Annick
? TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS
? TITLE OF INVENTION: GENES INVOLVED IN ARCHIDONIC ACID METABOLISM
? FILE REFERENCE: GENSET . 051CPI
? CURRENT APPLICATION NUMBER: US/09/641,638
? PRIOR APPLICATION NUMBER: US 09/502,330
? PRIOR FILING DATE: 2000-08-16
? PRIOR APPLICATION NUMBER: US 60/133,200
? PRIOR FILING DATE: 1999-05-07
? PRIOR APPLICATION NUMBER: US 09/275,267
? PRIOR FILING DATE: 1999-03-23
? PRIOR APPLICATION NUMBER: US 60/119,917

```

PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1123..3123
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3871..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5880
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5996..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon
LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc_feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTTTT
NAME/KEY: allele
LOCATION: 2048

OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G

NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTC
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match 2.0%; Score 65.4; DB 4; Length 20674;
Best Local Similarity 52.6%; Pred. No. 0.0004;
Matches 235; Conservative 0; Mismatches 206; Indels 6; Gaps 4;

1842 ACTGATGTCATATTTTCGCAATATTAATGAGTAATGCTACTGATCAATA 1901
11120 AATATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTA 11178
1902 GAAGAAACGCTGATATACATTTTAAAGACAAATTTTGAAGAAATGTATA 1961
11179 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11238
1962 TTTCATACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2021
11239 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11298
2022 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2081
11299 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11357
2082 TCTCATTAAT--ATTTTGAAATCTCAATTAATTAATTAATTAATTAATTA 2138
11358 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11417
2139 ATAGAAAGTCCCAAAAAATTTTGTACAGAAATTCCTCAATTTTGTATGAG 2197
11418 CTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11477
2198 AACCAAAATTAACAGATAGAAATCTATTTGTGGAATGGAAGTAGTAATTAATTA 2257
11478 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11537
2258 AGCAAAATTTAAAAAATTAATTAAGCC 2284
11538 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11564

RESULT 7
US-08-446-855A-1
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
TITLE OF INVENTION: phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573h Glebe Road, 8th floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match 2.0%; Score 63.6; DB 2; Length 8920;
Best Local Similarity 46.8%; Pred. No. 0.00074;
Matches 278; Conservative 0; Mismatches 304; Indels 12; Gaps 2;

1874 AATTAGATTCATGCTACTATCAATAGAGAAACGCTGAGTATTAATTAATTA 1933
152 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 211
1934 AAGACAAATTTTGAAGAAATGTTA-ATTCTACAAATTAATTAATTAATTAATTA 1992
212 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 271
1993 ATATGATATTCCTATGCTTAAATTAATTTTAAATTAATTAATTAATTAATTAATTA 2052
272 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 331
2053 GAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2112
332 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 391
2113 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2172
392 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 451
2173 ACTTCAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2221
452 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 511
2222 TATTTGTTGGAAGGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2281

LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc_feature
LOCATION: 17555..20674
OTHER INFORMATION: 3/regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903

OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match 2.0%; Score 63; DB 4; Length 20674;
Best Local Similarity 48.0%; Pred. No. 0.0012;
Matches 224; Conservative 0; Mismatches 230; Indels 13; Gaps 1;

```

OY 1825 TAATGACAGCCTGACAACTGTAGTCAATTTTCTGCAAAATTAATTAAGAAAT 1884
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11554 TAATTTAAATATTTCTTAGCTATTAATTAATTAATTAATTAATTAATTAAT 11495
OY 1885 CAATGCTACTATCAATAGAAACAGAGCTGATATACATTTTAATTTAAAGACAAAT 1944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11494 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11435
OY 1945 TTGCAAAATGTTAATTTCTACAAATATATTAATTAATTAATGATATATGATTC 2004
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11434 TTAATTTAATTTGACAGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 11375
OY 2005 CTATGCTCTTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAAT 2064
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11374 AATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11315
OY 2065 GTGGTGCAATTCAAATATC-----TCCATTAATTTTGTGAATCTACAAA 2111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11314 TTAAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11255
OY 2112 TTATTAATATTTAGTCAATTAACAATGATAGAAAGTCCAAAAAATTTGTTAAACA 2171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11254 ATATTAATAATTAATTAATTAATTTATTTAAATTAATTAATTAATTAATTAATTAAT 11195
OY 2172 AACTCCAAATTTTATTTTATTAAGCAACAATAACAGATAGCAAAATTTTGTG 2231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11194 TTAATTAATTTATTAATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 11135
OY 2232 TGCAATGCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11134 TTTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11088

```

RESULT 10

```

US-08-817-926-27
; Sequence 27, Application US/08817926
; Patent No. 6001590
; GENERAL INFORMATION:
; APPLICANT: Komeda, Toshinori
; APPLICANT: Komeda, Hisako
; APPLICANT: Tamai, Yukio
; APPLICANT: Iwamatsu, Akihiko
; APPLICANT: Kato, No. 6001590uo
; APPLICANT: Sakai, Yasuyoshi
; TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINI1
; TITLE OF INVENTION: FORMATE DEHYDROGENASE GENE
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,926
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/02597
; FILING DATE: 12-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 234133/1995
; FILING DATE: 12-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 42536/1996
; FILING DATE: 29-FEB-1996

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 081356/0112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4818 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Candida boidini1
US-08-817-926-27

```

```

Query Match 2.08; Score 62.6; DB 3; Length 4818;
Best Local Similarity 44.48; Pred. No. 0.001;
Matches 338; Conservative 0; Mismatches 419; Indels 4; Gaps 2;

```

```

OY 1525 ACTATTAATTTGCTTTCTTAATGCAATCTTAACATTAATTTAGTGGAAATTAAT 1584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4020 AATCTTTAATGATATTTCTTTATTTGATTTTTCATATATTTCTGATTAATTAATTTGTTG 4079
OY 1585 AAGAGCGAATCTGTAACATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4080 TTGAATTTCAATGATTAATTTATTTATTTCTTACAGATTTTAAAAAATTTTCCCTT-TTTT 4138
OY 1645 TACATACCTTTGGAGTTTAACTTAATCTTGTCTTCCTACGGTAAATTAATTAATCAT 1704
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4139 TATTTTCTTTGTTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4198
OY 1705 CGAGTAAAAAAGTTTGTCTTATTTCCGAGTCATGAAGGATTAACCTTAATGACTTT 1764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4199 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4258
OY 1765 AATTTTGGAAATGTAACCTTTTACT--CATGATTAATTAATCCGATGATTTTGTG 1821
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4259 CTGAGTTTAATGGAATTAACAGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4318
OY 1822 CCATTAAGCAGCCCTTCAACCTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1881
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4319 TTAATCATTTGGTATTTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4378
OY 1882 ATTCATGCTACTATCAATTAAGAAACAGCTGATTAATTAATTAATTAATTAATTAATTAATTAAT 1941
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4379 CTTTTGTGTTTATTTCTGTTTAAATTTTGTATCAATCTTTAAATTTAGTTTAA 4438
OY 1942 ATTTTGGAAAAATGTTAATTTCTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2001
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4439 ATTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4498
OY 2002 TTCCATGCTCTTAATAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2061
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4499 TTTCAATTTTGTGAACAAACGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4558
OY 2062 ATAGTGTGCAATTCATATCTCATTAATTTTGTGAATCTACAAATTAATTAATTAATTAATTAAT 2121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4559 TTTGTTAAATTTGATTAACATTTTGTATTAATTTTGTATCAATCTTTAAATTTAATTAATTAATTAAT 4618
OY 2122 TTAGCAATTAACAATGATAGAAAGTCCAAAAAATTTGTTAAGCAAGAACTCCAAA 2181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4619 TTAATTAATGAAAAAACCTTAAAAAGGACTTCGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4678
OY 2182 TTTTATTTTATTAATGAACAATAATTAACATAGAAAAATTAATTTGTTGTTGATGAATGA 2241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4679 TGACATATATTAAGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4738
OY 2242 GTAGTAATATATCAATTAAGCAAAATTTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 4739 TACTAGTCAAGCTAATTAATTTCTATTTTAAAG 4779

RESULT 11

US-07-867-106-2/c

Sequence 2, Application US/07867106
Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L

TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526r1s
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867,106

FILING DATE: 19920625

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU 87 7187

APPLICATION NUMBER: PCT/AU90/00530

FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Peeney, Joanne Longo

REGISTRATION NUMBER: 35,134

REFERENCE/DOCKET NUMBER: RICE-0002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 5852 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLCULE TYPE: DNA (genomic)

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

Query Match 1.9%; Score 61.8; DB 1; Length 5852;

Best Local Similarity 48.3%; Pred. No. 0.0015;

Matches 209; Conservative 0; Mismatches 217; Indels 7; Gaps 1;

Db 5607 AAAAAAAAAAAGTAAATTTGGAATTAATAAAGGCTTTT 5548

Db 2007 ATGCTCTAAATATTTTATA-----TTAGTATAATTCATTAGACCA 2059

Db 5547 AAAAAATATGATGATTTTATTTTAAATCATTGACGAGATTAATAA 5488

Db 2060 TAATAGTGTGATTCATTAATCTCCATTAATTTTGAATCTCAATTAAT 2119

Db 5487 TAAACATTTGATTTTATTTTATTTTATTTTAAATCAATTAATTAAT 5428

Db 2120 ATTTAGCATTAACATGCTAGAAAGTCCAAAAATTTGTTACGAACTTCA 2179

Db 5427 TAATATCTATTAATCTGATGATGAACTTAATTTAATTAATTAATTA 5368

Db 2180 AATTTTATTTT 2192

Db 5367 AATTTAGTATCT 5355

RESULT 12

US-09-316-083-2

Sequence 2, Application US/09316083A

Patent No. 6280942

GENERAL INFORMATION:

APPLICANT: The Institute of Physical and Chemical Research

TITLE OF INVENTION: Endonuclease

FILE REFERENCE: PH-651

CURRENT APPLICATION NUMBER: US/09/316,083A

CURRENT FILING DATE: 1999-05-20

EARLIER APPLICATION NUMBER: JP98/141861

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 2

LENGTH: 1431

TYPE: DNA

ORGANISM: Saccharomyces cerevisiae

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1428)

US-09-316-083-2

Query Match 1.9%; Score 61.2; DB 4; Length 1431;

Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

Db 1899 ATAGAGAAACAGCTGATATACATTTAATTAAGACAAATTTGAAAAATGTTA 1958

Db 164 ATATATTAATAAATTTTATTTATTTAGATATGATTAATGATTAATTAATCA 223

Db 1959 TAATTTCTAACAAATATTAATAAATATGAGCCATTAATGATTTCTATGTTTAAA 2018

Db 224 AACACGTACACCTGATTAATAAAGATTAAGAAATTAATAAATTAATGATTA 283

Db 2019 TATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2078

Db 284 AATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 343

Db 2079 ATATCTCAATTAATTTTGAATCTACA-AAATTAATTAATTTAGCAATTAACATG 2137

Db 344 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 403

Db 2138 CATGAGAGTCCAAAAAATTTGTTACAGAACTCCAAATTTTATTTATG 2197

Db 404 ATTTAAATTAATAAATCTTAATTAATTAATTAATTAATTAATTAATTAATG 463

Db 2198 AACAGAAATTAACAGATAGAAATATTTGTTGGAATGGAAGTAAATATACATTA 2257

Db 464 ATAAATTTTAATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 523

Db 2258 AGCAATTTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2317

Page 13

RESULT 15
 US-08-446-855A-1/c
 : Sequence 1, Application US/08446855A
 : Patent No. 5849573
 :
 : GENERAL INFORMATION:
 : APPLICANT: Stewart, Thomas S
 : APPLICANT: Flores, Maria V
 : APPLICANT: O'Sullivan, William J
 : TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
 : TITLE OF INVENTION: phosphate synthetase II
 : NUMBER OF SEQUENCES: 2
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Nixon & Vanderhye PC
 : STREET: 1100 No. 5849573th Glébe Road, 8th Floor
 : CITY: Arlington
 : STATE: Virginia
 : COUNTRY: USA
 : ZIP: 22201-4714
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn release #1.24
 :
 : CURRENT APPLICATION DATA:

[illegible]

Search completed: March 30, 2003, 09:01:35
Job time : 323.977 secs

GenCore version 5.1.4_P5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:25:26 ; Search time 243.877 Seconds

(without alignments)
11172.915 Million cell updates/sec

Title: US-09-502-426a-1_COPY_1_3202

Perfect score: 1 atgtggtattatattgtg.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_MNC

Gapop 10.0 , Gapext 1.0

Searched: 574371 segs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	2.5	640681	US-09-790-988-1	Sequence 1, Appl1
2	66.4	2.1	424	US-09-960-352-11218	Sequence 11218, A
3	65.2	2.0	335913	US-09-754-853A-2	Sequence 2, Appl1
4	65.2	2.0	335913	US-09-754-853A-3	Sequence 3, Appl1
5	64.4	2.0	473	US-09-960-352-11212	Sequence 1212, Ap
6	64.2	2.0	419	US-09-960-352-11234	Sequence 11234, A
7	64.2	2.0	640681	US-09-790-988-1	Sequence 2, Appl1
8	62.8	2.0	335913	US-09-754-853A-2	Sequence 1, Appl1
9	62.8	2.0	335913	US-09-754-853A-3	Sequence 3, Appl1
10	62.2	1.9	127197	US-09-754-853A-1	Sequence 1, Appl1
11	61.6	1.9	3410	US-10-196-063-1	Sequence 4756, Ap
12	61.4	1.9	1713	US-09-938-842A-756	Sequence 2, Appl1
13	61.2	1.9	1431	US-09-774-414-2	Sequence 154, Ap
14	61	1.9	302250	US-09-962-832-154	Sequence 3532, Ap
15	60.8	1.9	2000	US-09-938-842A-3532	Sequence 4, Appl1
16	60.4	1.9	513509	US-09-754-853A-4	Sequence 4582, Ap
17	60.2	1.9	393	US-09-960-352-4582	Sequence 4, Appl1
18	60	1.9	1864	US-10-105-481-4	Sequence 1, Appl1
19	59.8	1.9	3410	US-10-196-063-1	Sequence 1, Appl1

C 20	59.8	1.9	302250	10	US-09-962-832-154	Sequence 154, Ap
C 21	59.4	1.9	377	10	US-09-960-352-7419	Sequence 7419, Ap
C 22	59.4	1.9	5046	10	US-09-725-733A-13	Sequence 13, Ap
C 23	59	1.8	1872	9	US-09-938-842A-3504	Sequence 3504, Ap
C 24	58.6	1.8	53332	9	US-10-224-562-3	Sequence 3, Appl1
C 25	58.6	1.8	53332	9	US-09-801-861-3	Sequence 3, Appl1
C 26	58.4	1.8	516	10	US-09-960-352-5785	Sequence 5785, Ap
C 27	58.4	1.8	1109	9	US-09-938-842A-3533	Sequence 3533, Ap
C 28	58.4	1.8	3272	10	US-09-796-348-18	Sequence 18, Appl1
C 29	57.4	1.8	414	10	US-09-960-352-6528	Sequence 6528, Ap
C 30	57.4	1.8	513509	9	US-09-754-853A-4	Sequence 4, Appl1
C 31	57.2	1.8	2000	9	US-09-938-842A-3503	Sequence 3503, Ap
C 32	57	1.8	2000	9	US-09-774-414-2	Sequence 3059, Ap
C 33	56.2	1.8	1431	10	US-09-938-842A-3533	Sequence 2, Appl1
C 34	55.8	1.7	2000	9	US-09-938-842A-2991	Sequence 2991, Ap
C 35	55.6	1.7	53332	9	US-10-224-562-3	Sequence 3, Appl1
C 36	55.6	1.7	53332	10	US-09-801-861-3	Sequence 3, Appl1
C 37	55.2	1.7	740	9	US-09-791-279-27	Sequence 27, Appl1
C 38	55.2	1.7	2000	9	US-09-938-842A-2868	Sequence 13, Appl1
C 39	55.2	1.7	5046	10	US-09-725-735A-13	Sequence 6976, Ap
C 40	54.8	1.7	337	10	US-09-960-352-6576	Sequence 1, Appl1
C 41	54.8	1.7	684973	10	US-09-263-959-1	Sequence 4584, Ap
C 42	54.6	1.7	416	10	US-09-960-352-4584	Sequence 4527, Ap
C 43	54.2	1.7	1365	9	US-09-938-842A-4527	Sequence 11234, A
C 44	54	1.7	419	10	US-09-960-352-11234	Sequence 3, Appl1
C 45	54	1.7	6265	10	US-09-129-112-3	

ALIGNMENTS

RESULT 1
US-09-790-988-1
Sequence 1, Application US/09790988
Patent No. US20020127687A1
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: WATANABE, HIDEKI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 2.5%; Score 80; DB 10; Length 640681;
Best Local Similarity 45.8%; Pred. No. 0.0033;
Matches 395; Conservative 0; Mismatches 455; Indels 12; Gaps 3;

1559 AATATTATGTTAGTGGAAATTTAATGAAGGCACTTGTACATTAATTTATTTATTT
1619 AACTACTAGTATGTGATTTATCCAAATACATCTTGTGATTTAACTTAATCTGTTT
324746 AATCTTTTGTACAAAGAAATATAATATTTTATTTATCAATAATCCATTTTGTATC
1679 CTTCCAGCGTAAAT
324806 TCTTCATCATTTAAATAAATCAATTTCTTCTTAATATCAATATATCTTTTATATAT
1739 GCATGAGGATTAACCTATGACTTAATTTTGAATAATGATACCTTTTACATATAGA
324866 TGTGGGTGATTTTATGTCATTAATAATCTTAATAATTTTAAAGATTTTCTTCT

OY 1799 TTAATTACCGATGTTGGTGTGGTCATCATGCAGCCCTCACAATCGTAGTGCAATTT 1858
 ||| | | | | | | | | |
Db 324926 GTCAAAGAATAATTATTTTTCCGAAAATGAATATTAATBGGACAATACTGATTAITT 324985

OY 1859 TTTCGCCAAATATTAATAATTAGGAATTCAGTCTACTATCAATAGAAGAACGCTGACTA 1918
 ||| | | | | | | | | |
Db 324986 TTTTGTGATTAATTACATTTATTAATAATAGACTGATTAATTTTGATTTATTAATTAATA 325045

OY 1919 TTACATTTTAATTATTAAGCACAAAAATTTTGGAAAAATGTATTAATTTCTAACAAATTTATT 1978
 ||| | | | | | | | | |
Db 325046 ATTTTTTGGTTTTTAAAAATTTTATTTTGTAAATTTCCAAGAACGATATTCCTTTAAT 325105

OY 1979 AAAATATGATCGCCTATTAATGATTTCCCTA-----TGTCCTAAAAATTTTDTTATAT 2033
 ||| | | | | | | | | |
Db 325106 AAAGTAATRAATTTTTTAATTATTAATTAACTAATTTTATTTCARPTATATTTTTTTATPA 325165

OY 2034 TTAGTTATTAATPACATTTATGAAACCATAATATAGTGGTAATCCAATATCTCCATTAATA 2093
 ||| | | | | | | | | |
Db 325166 ATTAAAAATTTCTTAATAATAATATTTTAAATGTATTTAAAAAATAATAATATA 325225

OY 2094 TTTTGTGAATCTCAAAAT-----TATTAATATTGTAGCAATPACAATGCATAGAAAGTT 2148
 ||| | | | | | | | | |
Db 325226 ATTTATTTTATTTTACATTTTATTAATTAATTAATTAATTAATCTAATAATATTTTATATTTTA 325285

OY 2149 CCAAAAAAAAAATTTGTCTPACAGAAACTCCAAATTTTTTTTTTTATGACAAAGAAATA 2208
 |||| | | | | | | | | | |
Db 325286 AITMAAAATRCAATTTTATATTAATTAATAATACIGATARTTTTTCTTTTATTTTTTTAAATTT 325345

OY 2209 ACAGATAGAAAACTATTTTTGTGTGGAAATGGAAGTAGTAATATACATTTAGCAAAATTTTA 2268
 ||| | | | | | | | | |
Db 325346 ACTATTAATGAAT- -ATTTAGTATTTATTAATAAATPAGTAATAATAAGNACAAAAAATA 325403

OY 2269 AAAAATTTATPAGGCTATAGGCGCTCAAGATAGTTATCTAGTAGCGTATTAATTAAT 2328
 ||| | | | | | | | | |
Db 325404 TCANATATAAATAATPAGGAAGTATTTCTTTTTTTTATCTAATTTTATATTAATAAAAA 325463

OY 2329 GCATGGTCGATTCGAAATTTGGGACACACATGAAACCGAATTAATAATTTTAACTTTTAA 2388
 ||| | | | | | | | | |
Db 325464 GAATTTGCATGGAATCGATTAATAATTAATAATTTTTTATTTTCAATTTGAATATTTCAAACA 325523

OY 2389 ATAAATTAATAATTTGAGTAAAT 2410
 ||| | | | | | | | | |
Db 325524 TTCATTAATAAACTTTATATAATTT 325545

```

RESULT 2
US-09-960-352-11218/C
Sequence 11218, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathalagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 11218
LENGTH: 424
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 48-LIB3058-026-Q1-K1-D12
US-09-960-352-11218

Query Match          2.1%; Score 66.4; DB 10; Length 424;
Best Local Similarity 48.4%; Pred. No. 0.038;
Matches 184; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

```

Db 396 TATATATTTTTTCCCAATATATATATCCCTCATATTTTTCCTTTTTATATTCCTTTAT 339

Oy 68 TTACATGATTTGAGTTGGTTCCTCAATTTGGATTCACAGATATTAATATTAATAATTC 127

Db 338 TTTTAAATTTTAAATCTTTTTTTTTTTTAAATATAGCTATATCTTTTTAAATTTTCAT 279

Oy 128 ATTTAAATATTTACAGATATTAATTAATCTTTACATGTATGTATTAACAAATATCT 187

Db 278 TTTTTTTTTTCTTAAATAATCTCTTTTAAATTTTTTTTTTAAATTCCTATATATATTTT 219

Oy 188 ATCTTGGATATGAGAAATATGAGTTTGGATTTATATATATTAAGAAATATTCGA 247

Db 218 TTTTATATTTTAAACAAATATTTAAATTTATATATTTATTTTAAATTTTTTTTA 159

Oy 248 TTCCATTTGGTGAATACAGTTAAGTTTTTGTCTTCTTTGTTATATGTATATAGAG 307

Db 158 TTTAA 99

Oy 308 TAAATCAAAAAGAGTTGATGAAAGTAAACATATTTTCGTATGACCCCAAAAAAAA 367

Db 98 AA 39

Oy 368 AAAAAAAAAACAACAACA 387

Db 38 AAAAAAAAAAAAAAAAAAAAA 19

```

RESULT 3
US-09-754-853A-2
: Sequence 2, Application US/09754853A
: Publication No. US20030005491A1
: GENERAL INFORMATION:
: APPLICANT: Hauge, Brian M.
: APPLICANT: Parnell, Laurence D.
: APPLICANT: Parsons, Jeremy D.
: APPLICANT: Wang, Ming Li
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: FILE OF INVENTION: Soybean Cyst Nematode Resistance
: TITLE REFERENCE: 38-10(15810)B
: CURRENT APPLICATION NUMBER: US/09/754,853A
: CURRENT FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 60/174,880
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 1119
: SEQ ID NO 2
: LENGTH: 335913
: TYPE: DNA
: ORGANISM: Glycine max
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (45163)..(45314),(45450)..(45509),(46941)..(48763),(48975)..(49573)
: OTHER INFORMATION: Clone ID: 24001_region_G3
: US-09-754-853A-2

```

Query Match	2.0%;	Score 65.2;	DB 9;	Length 335913;
Best Local Similarity	46.1%;	Pred. No. 0.67;		
Matches	399;	Conservative	0;	Mismatches 459; Indels 9; Gaps 5
QY 1582	AATAAGACCGA	ACTGTGTACATTTACAAATATTATTTTGATGACTGATGCTATTATGCC	1641	
Db 279575	ATTAAATATAAAAAATATTTAAATATTAACATTTTATAGAAAAATATATACAGTAAATATAT		279634	
QY 1642	AAATCATACCTTTGGATCTTTAAACCTTAATCTGTTCTTCCCTACGCTATAAATATTAT		1701	
Db 279635	TTTAAATCAAAAATCTCTCTTTATTTTATTTTATTTTATTTTAAAGACAAAAGATTAAACAT		279694	
QY 1702	CATGAGGTAAAAAAGTTTGTCTATTTCGCAATGCATGCAAGATTAACCTAATGAC		1761	
Db 279695	AACGTATTTCAAAATAGCTTAGCTTTGATTTTACGACAGTGGAAATCGGATAGGACCCA		279754	
QY 1762	TTTATTTTTGAAAATGTAAACCTTTTACATCATGATTAATTTACCGTATGTTTGTGTG		1821	

Query Match	2.08;	Score 65.2;	DB 9;	Length 335913;
Best Local Similarity	46.18;	Pred. No. 0.67;		
Matches 399;	Conservative 0;	Mismatches 458;	Indels 9;	Gaps 5;
QY 1582	AATAAGACGGAACCTGGTACATCAATATTTATATAGATACAGATGTGATTTGCC	1641		
Db 279575	ATTATATATATAAATATTAATTAATTTATACAAAAATTAACAGTAAATATTAT	279634		
QY 1642	AAATGCAATCTTGGAGTGTAAACCTTAATCTGTGTTCTTCCACGATATAATATTAT	1701		
Db 279635	TTTATATCAAAAAATTTCCCTTTTATTTTTTATAGTATATTTTTTAAGCAAAAGATTAACT	279694		
QY 1702	CATCGAGGTAAAAAAAGTTTGTGCTATTTTCGGAGATGCATGAGAGATAAACCTAATGAC	1761		
Db 279695	AACGATTTTCAAAATATACCTTAGCTTAGCTTATTTTACGACAGTGAATCGGGTAAGGCCCA	279754		
QY 1762	TTTTATTTTTTTGAAAAATNGTAAACCTTTTACTCATATTAATTAATTCGGATGTTTTGGTG	1821		
Db 279755	ACTATATCTTTT-TTGGCTTTTTCATTTTTCATTTTCATTCGCGTTTTTCTTTTCATTTATGTTTC	279813		
QY 1822	CCATATGACAGCCCTACAACTGTGATGTGCAATTTTTTTCGCAATTTTAATTATAGGA	1881		
Db 279814	CTTATATTAATATATATAAAGCTACAAAGTGTGTTGTATATATGAGAAATATGAAACCAACA	279873		
QY 1882	ATTCAATGCTACATCAATATGAAAGAAACAGCTGATATTCATTTTAATTAATTAAG	1940		
Db 279874	TAAAGAAAAAGTATATGATTAAGAAACCTAATTAATACCTAATCTATATATTTGATGCCGAT	279933		
QY 1941	AAATTTTGGAAAAATGTATTAATTTTCAACAATATATATTAATTAATTAATGATCCATATAGTA	2000		
Db 279934	AAATATTCGATACAGATATATATTTTCAACATCTGCTTAATTAATTAATTTTGGTAAAAAT	279993		
QY 2001	TTTCCATATGTTCTTAAATATTTTTTTTTTATATTTAGTATTAATTAATCAATATGAAACCAAT	2060		
Db 279994	ATATTAATACCTTATATATAAAATTAATTAATCTTTTAATTAATTTTATATA-----ATTATAAAAAGAA	280049		
QY 2061	AATAGTTGGGAATTCGAATATCTCCATTAATTTTTTGAATCTACAAATATTATATA	2120		
Db 280050	ATTATATTTTAAACAAATATTTAGCCAAATGTATTTTATTTATGCGATTACTTTTGTATA	280109		
QY 2121	TTTATGCAATTAACAATGCATAGAGAAATGCCAAAAAAATTTTGTATACAGAAACTGCCAA	2180		
Db 280110	--AAATTAATAACAAATTAACAATTAATGCTGCAAAATGTTTTTCTAAATATTTCAATATA	280167		
QY 2181	ATTTTTTTTTTTTAAGCAACAAATTAACGATATGAAATCTATTTTGTGTGGAATGGA	2240		
Db 280168	ATTATTTATTTTAAAAATTTTAATTAACAATTTTAATTTTAATTTTAATTTAATGATGC	280227		
QY 2241	ACTAGTAATATACATTAAGCAAAATTTTAAAAAATTAATTAAGCTATACGGCGCAAGT	2300		
Db 280228	AAAACATATTTATTAATTAATAGCGTGATTAATAAGATATATATATATATATGTTATATTA	280287		
QY 2301	ATGTATATCTAGAGGTGTAATTAATTAATGCAATGCGGATTCAGAAATGGGACACAAATG	2360		
Db 280288	ATTAGTATATAGATATATTAATTAATTAATTAATA-TAAATCTCAATTTTCCAAATAATA	280346		
QY 2361	AAAACGAATTAATAATTTAACTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2420		
Db 280347	TTATACATATTTCTATGTAATGAGCTAAATATTATTAATTAATTAATTAATTAATTAATTAATG	280406		
QY 2421	ACTATTTGGGGGCAAAAAAAGACAA 2446			
Db 280407	TTTATTTACCTGTTTAACTAATCAAGA 280432			

RESULT 8
 US-09-754-853A--2/c
 Sequence 2, Application US/09754853A
 Publication NO. US20030005491A1
 GENERAL INFORMATION:
 APPLICANT: Hauge, Brian M.
 APPLICANT: Parnell, Laurence D.
 APPLICANT: Parsons, Jeremy D.
 APPLICANT: Wang, Ming Li
 TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
 TITLE OF INVENTION: Soybean Cyst Nematode Resistance
 FILE REFERENCE: 38-10(15810)B
 CURRENT APPLICATION NUMBER: US/09/754,853A
 CURRENT FILING DATE: 2001-01-05
 PRIOR APPLICATION NUMBER: US 60/174,880
 PRIOR FILING DATE: 2000-01-07
 NUMBER OF SEQ ID NOS: 1119
 SEQ ID NO 2
 LENGTH: 335913
 TYPE: DNA
 ORGANISM: Glycine max
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (45163)..(45314)..(45450)..(45509)..(46941)..(48763)..(48975)..(49573)
 OTHER INFORMATION: Clone ID: 240017-region_G3
 US-09-754-853A--2

QY	2103	ANCTCAAAATTTTAAATTTTGGTCAATACATGAGAAAGTCCAAAAAAATTTT	2162
Db	280254	ATACAGCTTAATTAATAATAAGTTTGATACATTAATTTAAATTAATAATTTGT	280195
QY	2163	GTTAACAGAACTCCAAATTTTTTTTGTATGAGCAAAATAATACAGTACAAAAT	2222
Db	280194	TAAATAAATTTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAAT	280135
QY	2223	ATT--TTGTGTGGCAATGGAAGTAGAATATACATTAACAAATTTAAAAATTAATA	2280
Db	280134	AATAATTTGTAATTTGTTATATATTTTAAACAAAAGTATGCATTAATTAATTAAT	280075
QY	2281	AGCCATACGGGCTCAAAAGTATGTACTAGTAGTAATTAATAATGATGGTGGAT	2340
Db	280074	GGCATAATTTTGTTTAAAAATTAATTTCTTTTTPAAATTAATAATAATTAATAAGTT	280015
QY	2341	TCAGAAATGGACAACAATGAAAAAGCAATTAATACTTTAAATTAATAATAAAAT	2400
Db	280014	-----AATTTTATTAAGAATTAATATATTTTAAACGAATAATATTTATTAACCATAGTT	279599
QY	2401	TTGAGTAATAGTGTTTTCTGACTTTGAGGGGCAAAAAAGACAATGCCAAAAGTCTAC	2460
Db	279958	GAAATTAATCTGATCGATATTTTATACGATCCAAATAATATATAGATTAAGGTTTATT	279899
QY	2461	GGGTTT 2466	
Db	279898	AGGTTT 279893	

```

RESULT 9
US-09-754-853A-3/c
: Sequence 3, Application US/09754853A
: Publication No. US20030005491A1
GENERAL INFORMATION:
APPLICANT: Hauge, Brian M.
APPLICANT: Parnell, Laurence D.
APPLICANT: Parsons, Jeremy D.
APPLICANT: Wang, Ming Li
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: TITLE OF INVENTION: Soybean Cyst Nematode Resistance
FILE REFERENCE: 38-10(15810)B
CURRENT APPLICATION NUMBER: US/09/754, 853A
CURRENT FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: US 60/174, 880
NUMBER OF SEQ ID NOS: 1119
SEQ ID NO 3
LENGTH: 335913
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS
LOCATION: (46798)..(48763)..(48975)..(49573)
OTHER INFORMATION: Clone ID: 240017-reg1on_c3
US-09-754-853A-3

```

[illegible]

OY	2103	ATCTCAAAATTTATTAATTTTATAGCAATTAACAAATGCATPAGAAAGTCCAAAAAAATTTT	2162
Db	280254	ATCAGCCATATTTATTTAATAATAGTTTGCATGCAATAAATTTTAAATTAATAATTTAAATGT	280195
OY	2163	GTTAAACAGAAATCTCCAAATTTTATTTTATGGAACAGAAATAACATAGAAAACT	2222
Db	280194	TAAATTAATTTTAAAAATTAATAATTAATTAATGAAATATTGAAAAACATATTTCAGC	280135
OY	2223	ATT--TTCTTGTGCAATGGAAAGTGTATATATACCTTTAAGCAAAATTTAAAAAATTATATA	2280
Db	280134	AATATATGTAAATTTGTTATATATTTTAAACAAAAGTATGCAATTAATAATATACATTT	280075
OY	2281	AGCCATACGGCGTCAAAAGTATGTTATCTAGTACGTAAATTAATATGATGGTGGAT	2340
Db	280074	GGCATTAATTTGTTAAAAATTAATTTCTTTTATTAATTTAATAAATTTAAATTTAAAGTT	280015
OY	2341	TCAGATATGGGACACAATGAAAAAGAAATTAATAATTTAACTTAATAATTAATAAAT	2400
Db	280014	----AATTTTATAAAGTATATATATATTTTAAACGAAATATATATTTTATTTAACCATATGTT	279955
OY	2401	TTGAGTAAATGTGTTTCTGACTATGAGGGGCAAAAAAAGACAAATGCCAAAAGTCATC	2460
Db	279958	GAAATTAATCTGATGTCGATTTTATTAACGCATCAAAATTAATATAGATTAGATTATTT	279899
OY	2461	GGGTTT 2466	
Db	279898	AGGTTT 279893	

```

RESULT 10
US-09-754-853A-1
: Sequence 1, Application US/09754853A
: Publication No. US20030005491A1
: GENERAL INFORMATION:
:   APPLICANT: Hauge, Brian M.
:   APPLICANT: Parnell, Laurence D.
:   APPLICANT: Parsons, Jeremy D.
:   APPLICANT: Wang, Ming Li
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: FILE OF INVENTION: Soybean Cyst Nematode Resistance
: TITLE REFERENCE: 38-10(15810)B
: CURRENT APPLICATION NUMBER: US/09/754, 853A
: CURRENT FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 60/174, 880
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 1119
: SEQ ID NO 1
:   LENGTH: 127197
:   TYPE: DNA
:   ORGANISM: Glycine max
: FEATURE:
: OTHER INFORMATION: Clone ID: 515002_region_G2
US-09-754-853A-1

```

Query Match	1.9%	Score 62.2;	DB 9;	length 127197;
Best Local Similarity	43.5%;	Pred. 0.1,5;		
Matches	383;	Conservative	0;	Mismatches 493; Indels 5; Gaps 2
QY 1371	ATCACTGCAAAATATCATCTCTGGTGGTCACAAATPAAAAACAAAAACAAGAAAAAGAA	1430		
Db 71604	ATCATTAATATTCACAAATTCATCTTTATATPAAATATATATACAAAAATCTAATATTATCTA	71663		
QY 1431	AACGATTTTCTTGATTCATCAATGATCTPAAAAATGCATGATCTTTGGGTACACT	1490		
Db 71664	AA--ATATTTATATTTTGGTTATATATGATTTAAATATGATCTATATATGAAATTTTAAAA	71721		
QY 1491	TTGCAAGTCTCTACAAACGCTGAACATCTGCAACTATPAAATGCTTCTTTAATGCA	1550		
Db 71722	TATAAATATTTGATATATTTTTTTATTTGTTGTCATTTTAAATAAATTTGATTTCAATGA	71781		
QY 1551	TCTTTAACATCTTTTCTTTAGT---GGAAATTAATGAACGGAACCTGTGAACATTTCA	1607		
Db 71782	TAGTATATATATATATGATATATTTAAACGATTAGATTTAAATAAAAAATATATTTATATAT	71841		

QY	1608	AAATTTATTTTGGAACTAGTATGATTTATTCGAAATTCAGATCTGGATGCTTTAAACT	1667
Db	71842	ATATATATATATATATAAAAATTTATTTGTATATATTTAATCGACATTTGGCGTATATATAG	71901
QY	1668	TAACTTGTTCCTCCACGATATAAATATATATCATCGAGGTAAAAAAGTTTGTCTT	1727
Db	71902	TCTGTGTTTTCATGTTGGATTTCTTCTTCGCAACCATATATAGCACTATATAGG	71961
QY	1728	ATTTCGGCATGCAAGGAATTAACCTATATACCTTTAATTTTGGAAATGTAAACCCTT	1787
Db	71962	TTTTGTAAATGTAACATCGGGCAACAACCTGTGCTCTTATTTCTTTAAATAAATTTGATCCA	72031
QY	1788	TTATCATAGATATTAATTAACCTATGTTTTTTGTGGCATATATGACAGCCCTCAACTGTG	1847
Db	72022	TTAGCTGTTTTTTAAAAAAAATTTGTGGAGATATAGTCGATATCAATTTTCTTATATA	72081
QY	1848	ATATGTCAAATTTTTCGCAATATTTAAATTAGAATTCAGATCTACTATCAATAGAACAA	1907
Db	72082	AAATATGATATCAACGACATATAAAAAATATGAGACGTAAATTAACAATATCTTATGCGA	72141
QY	1908	ACAGCTGATATTAATTTATTTAATGAAGACAAATTTTGAAGAAATGTTATATTTCTTA	1967
Db	72142	AATGATTAACGATATCTTAAAAAGTTTATTTAAAAACTTAAAGTAAAAATATTTTATATA	72201
QY	1968	ACAATATTTATTAATAATATGATGCCATATATGATTTTCCATGTCTTAAAAATTTTCTT	2027
Db	72202	GAAAGAAAAAATATATATTTATTTACAAATATTTTTTTATATCATCTTTTGAATTTCAAAA	72261
QY	2028	TTATTTTATAGTTATTAATATACATTATGAACCAATATATAGTGTGCAATTCAAATATCTCCA	2087
Db	72262	TAAATTTTAAATTTCTTAAACACATACATGATGTAATGCAAAATCCATATATATATATATA	72321
QY	2088	TTAATATTTTGAATCTACAAATTTATTAATTTAGCATATTAACAATGCATAGAAAGT	2147
Db	72322	TA	72381
QY	2148	TTCAAAAAAAATTTTGTATACAGAACTTCCAAATTTTTTTTTTTATATGAACAAGAAAT	2207
Db	72382	TATGATTAAGAACCTCTCCCTACTACATCCCATCTATGCATAGTGTGTTGCTATACGCCACTT	72441
QY	2208	AACAGATGAAAAACATATTTTGTGTGGAAATGGAATAGTAA 2248	
Db	72442	ACCACCAATCTCAAGTGTCTGCACTAAGTCCAAATTAATA 72482	

```

RESULT 11
US-10-196-063-1
: Sequence 1, Application US/10196063
: Publication No. US20030027192A1
: GENERAL INFORMATION:
: APPLICANT: Gorovsky, Martin A.
: APPLICANT: Shang, Yuhua
: APPLICANT: Song, Xiaoyuan
: TITLE OF INVENTION: TETRAHYMENA METALLOTHIONEIN GENE PROMOTER AND ITS USE
: FILE REFERENCE: 176/61102
: CURRENT APPLICATION NUMBER: US/10/196, 063
: CURRENT FILING DATE: 2002-07-15
: PRIOR APPLICATION NUMBER: 60/305,167
: PRIOR FILING DATE: 2001-07-13
: PRIOR APPLICATION NUMBER: 60/317,322
: PRIOR FILING DATE: 2001-09-05
: NUMBER OF SEQ ID NOS: 1
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 3410
: TYPE: DNA
: ORGANISM: Tetrahymena thermophila
US-10-196-063-1

```

Query Match	1.98;	Score 61.6;	DB 9;	Length 3410;
Best Local Similarity	45.48;	Pred. No. 0.49;		
Matches 305;	Conservative 0;	Mismatches 359;	Indels 8;	Gaps 2

```

QY 1550 ATCTTACATATTTATTTAGTTGGAATTTAATAAGCCGAACCTTGACATTAACAT 1609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 569 AGCTTAAATTAATGATGCTTTCATCCCTTAAATGAGATCTGCTATTAATTCATTTCTGT 628
QY 1610 ATTTATATATGATGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 1669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 TTACCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 681
QY 1670 ATCTGTTCTTCTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 682 TTGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 741
QY 1730 TTTCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 742 TTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 801
QY 1790 ACTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 802 AATTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 861
QY 1850 AGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 862 TTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 921
QY 1910 AGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 922 AATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 981
QY 1970 AATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 982 ATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1041
QY 2030 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1042 TTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1101
QY 2089 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1102 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1161
QY 2149 CCAAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 TTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1221
QY 2209 ACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1222 CATTTAAGTAAA 1233

```

RESULT 12

```

US-09-938-842A-4756
; Sequence 4756, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4756
; LENGTH: 1713

```

```

; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-4756

```

```

Query Match 1.9%; Score 61.4; DB 9; Length 1713;
Best Local Similarity 46.6%; Pred. No. 0.41;
Matches 229; Conservative 0; Mismatches 261; Indels 1; Gaps 1;

```

```

QY 1928 AATTAAAGCAATTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 AATTAAATTAATTAACCTATTTTAACTGACATGATGATGATGATGATGATGATGAT 710
QY 1988 TGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2047
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 AATGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 770
QY 2048 AATATGACCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 AATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 830
QY 2108 CAATTAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 889
QY 2168 CAGAACTGCAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 890 TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 949
QY 2228 GTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 950 AATGTAAGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1009
QY 2288 ACGCGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1010 AATTTGTTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1069
QY 2348 TGGGACACAGCAAGAAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1070 ATGAATGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1129
QY 2408 AATGCTTTTC 2418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1130 ACTTTGATTC 1140

```

RESULT 13

```

US-09-774-414-2
; Sequence 2, Application US/09774414
; Patent No. US20020102231A1
; GENERAL INFORMATION:
; APPLICANT: The Institute of Physical and Chemical Research
; TITLE OF INVENTION: Endonuclease
; FILE REFERENCE: PH-651
; CURRENT APPLICATION NUMBER: US/09/774,414
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 09/306,970
; PRIOR FILING DATE: 1999-05-07
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1428)
US-09-774-414-2

```

```

Query Match 1.9%; Score 61.2; DB 10; Length 1431;
Best Local Similarity 47.4%; Pred. No. 0.41;
Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;
QY 1899 ATAGAAAGAAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1958

```

D	b		164	A T A T A A T A A A A A T T T A T A T A T A T A T A G A T A T G A T A T A T A T A T A T A T A T A T A A C T A C A T A	223
O	y		1359	T A A T T T C T A A C A A T A T T A T T A A A A T A T G A T G C C T A T A T G T A T T T C T A T G T T C T T A A A	2018
D	b		224	A A C A G C T A C C A C T A T G C T A A T A A A A A G A T T A T G A T T C A A A A A A T A T T A T A G T A T A	283
O	y		2019	T A T T T T T T T T A T A T T A G T A T A A T T A C A T A T T A T G A C A A T A A T A G T T G G A T T C A A	2078
D	b		284	A A T T A T T A T A C T T A T T T T T A T A T T T T A A A T A A A A T G A A A T G A A A T G C A A T A T A T A	343
O	y		2079	A T A T C C C A T A A T A T T T T T T T G A A A C T F C A - A A T A T T A A T A T T A G C A A T A A C A T G	2137
D	b		344	A T A T A T A T A A T A A T A T A T A T T T C A T T A A A T A T A T A T A A T T A A A A A T A T A T A G A A T A	403
O	y		2138	C A T A G A A G A T T C C A A A A A A A A T T T G T A A C A G A A C T T C C A A A T T T T T T T T T A T A G	2197
D	b		404	A T T A A A T T A A A A C T C T A A A T A T T G A A C A C T A T T A C T A A T A T T T T A T T A A T G	463
O	y		2198	A A C A G A A A T A C A G A T A G A A A A C T A T T T G T T G G A A T G A A G A C T A G T A A T A C A T T A	2257
D	b		464	A T A A A T A T T A A T T A A A A A A A A T A T A T A G A A A T T A G A T A T T A A T A A T A T A G A T T C C T A A T A	523
O	y		2258	A G C A A T T T A A A A A A T T A T A T A G C C T A T A G C G C T C A A A G T A G T T A T C T A G A G T G	2317
D	b		524	A T T A T A G T T A A T A A T A A T T A A T A T A T A A A G T A A A T A A T A A T T T A A A C A C A T A T A G A T T	583
O	y		2318	T A A T A T A A T A A T G A T G C G A T T C A G A A T T G G	2351
D	b		584	T A A A T A A T A A T G A T T T A T A G A T A T T T A T A C A G	617

```

RESULT 14
US-09-962-832-154
: Sequence 154, Application US/09962832
: Patent No. US20020110821A1
: GENERAL INFORMATION:
: APPLICANT: Ebner, Reinhard
: TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signature
: TITLE OF INVENTION: Sets
: FILE REFERENCE: 689290-74
: CURRENT APPLICATION NUMBER: US/09/962,832
: CURRENT FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: US/60/235,077
: PRIOR FILING DATE: 2000-09-25
: PRIOR APPLICATION NUMBER: US/60/235,280
: PRIOR FILING DATE: 2000-09-25
: NUMBER OF SEQ ID NOS: 259
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 154
: LENGTH: 302250
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-962-832-154

```

[illegible]

Dd 300362 AATATAATTAAATAAATAAATAATTAATAATTAATAATTTTATTTTAAAAATAATATTT 300361
Oy 2180 AATTTTTTTTTTTATGCAACAACAAAATAGACATATGGTGCGAATGG 2239
Dd 300362 ATATTATATATTAATATTAATTAATAAATAATTAATGAACCTTATATTAATATATTAT 300421
Oy 2240 AACTAGTAATATACATTAGCAATTTTAAAAATTTATAT 2280
Dd 300422 ATATTATATATATATATATTAATTAATAGATATTAATAAAAA 300462

```

      RESULT 15
      US-09-938-842A-3532/C
      ; Sequence 3532, Application US/09938842A
      ; Patent No. US20020160378A1
      ; GENERAL INFORMATION:
      ; APPLICANT: Harper, Jeff
      ; APPLICANT: Kreps, Joel
      ; APPLICANT: Wang, Xun
      ; APPLICANT: Zhu, Tong
      ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
      ; TITLE OF INVENTION: SAME, AND METHODS OF USE
      ; FILE REFERENCE: SCRIPT300-3
      ; CURRENT APPLICATION NUMBER: US/09/938, 842A
      ; CURRENT FILING DATE: 2001-08-24
      ; PRIOR APPLICATION NUMBER: US 60/227,866
      ; PRIOR FILING DATE: 2000-08-24
      ; PRIOR APPLICATION NUMBER: US 60/264,647
      ; PRIOR FILING DATE: 2001-01-16
      ; PRIOR APPLICATION NUMBER: US 60/300,111
      ; PRIOR FILING DATE: 2001-06-22
      ; NUMBER OF SEQ ID NOS: 5379
      ; SEQ ID NO 3532
      ; LENGTH: 2000
      ; TYPE: DNA
      ; ORGANISM: Arabidopsis thaliana
      US-09-938-842A-3532
  
```

	Query Match	1.9%	Score 60.8	DB 9	length 2000
	Best Local Similarity	47.0%	Pred. No. 0.54		
	Matches 188	Conservative 0	Mismatches 212	Indels 0	Gaps
QY	1950	AAAAATGTAATATTTCTACAAATATTTAAAAATATGATGCTATATATGATATTTCCATATG	2009		
DB	811	AAAAATGACAGAACCAAGAGAGCAAAAAAATTAATATTCGTTATTTATTAATCAAA	752		
QY	2010	TTCTTAATAATTTTTTTTTTATATATAGTATATAATACATTATNGAACCAATATAGTGG	2065		
DB	751	TTTAAATATGCTTATATATATATTTATATTTCTATATATTAATATTTCTATATCTTTAG	692		
QY	2070	TGAATTCAAATATCTCCATTAATATTTTTTGAATCTACAAATTTATATATTTAGTCAA	2139		
DB	691	TATATATGAT	632		
QY	2130	TACAAATGCAATGAAAGTTCCTCAAAAAAATTTTGTTACAGAACTTCCAAATTTTTTT	2188		
DB	631	TATGATATTTATTTTAAATTTAGTATATAAATATCTTTTTTAAAGGAATTTTATTTAAATTT	572		
QY	2190	TTTTATGGACAGAAATTAACATATGAAATCAATTTTGTGCAATGAGAGTATAT	2249		
DB	571	GTAATATATATTTAAATTTTACATCTATATAAATATATATTTTGAAGGAAATTTATAT	512		
QY	2250	ATACATTTACGAAATTTTAAAAAATTTATATATACCTATATACGGCTCAAAAGTATGTAATCT	2305		
DB	511	TTAATTTATTTTAAATTTCAATACAGAACTTTAATTTAAATGAAATTAATTAACAAATTC	452		
QY	2310	AGTAGGTATATTAATAATGCAATGGCGCATTCAGATATG	2349		
DB	451	AATAAATGATTTTTTTAGTATATCTTATATGTTAAATATTTAG	412		

Search completed: March 30, 2003, 13:11:02
Job time : 4203.88 secs

Mon Mar 31 09:27:33 2003

us-09-502-426a-1_copy_1_3202.rnpb

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:08:52 ; Search time 4178.19 seconds

(without alignments)
12411.569 Million cell updates/sec

Title: US-09-502-426a-1_COPY_1_3202

Perfect score: 3202

Sequence: 1 atgttggtatcatattgttgc.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estb:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estow:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vtl:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195	6.1	473	17	AL768954
2	148	4.6	148	17	BH811044
3	146.4	4.6	473	17	AL768954
4	113.2	3.5	1101	17	CNS00EVL
5	110.8	3.5	1101	17	CNS00EVL
6	109.6	3.4	1101	17	CNS0021T

C	7	98.4	3.1	836	17	CNS01100
C	8	98.2	3.1	1101	17	CNS00E07
C	9	98	3.1	807	10	AV717372
C	10	1092	3.1	1092	17	CNS020K7
C	11	94	2.9	1101	17	CNS003BD
C	12	94	2.9	1201	17	CNS0167M
C	13	93.6	2.9	1101	17	CNS003BD
C	14	93.6	2.9	1101	17	CNS00E07
C	15	92	2.9	661	17	CNS00B01
C	16	91.8	2.9	886	17	CNS020Y7
C	17	91.8	2.9	886	17	CNS070X
C	18	91.8	2.9	886	17	CNS0039G
C	19	91.8	2.9	1101	17	CNS04D0K
C	20	90.6	2.8	945	17	CNS01100
C	21	90.6	2.8	1187	17	B11102
C	22	90.4	2.8	836	17	CNS01100
C	23	89.6	2.8	1029	17	CNS012GM
C	24	89	2.8	987	17	CNS014PQ
C	25	88.6	2.8	1092	17	CNS020K7
C	26	88.6	2.8	1190	17	CNS020M7
C	27	88.4	2.8	1190	17	CNS0161D
C	28	88.2	2.8	1101	17	CNS0039G
C	29	87.4	2.7	1169	17	CNS06RHO
C	30	87	2.7	1225	17	CNS0161D
C	31	86.8	2.7	1101	17	CNS0021J
C	32	86.6	2.7	1101	17	AO258984
C	33	86	2.7	1013	17	CNS06RHO
C	34	85.8	2.7	1187	17	B11102
C	35	85.6	2.7	961	17	CNS008H1
C	36	85.2	2.7	928	17	CNS00DKY
C	37	85	2.7	1101	17	CNS05AB2
C	38	84.8	2.6	1084	17	CNS071NH
C	39	84.8	2.6	1101	17	CNS003BH
C	40	84.6	2.6	1190	17	CNS020N7
C	41	84.4	2.6	804	17	B12681
C	42	84.4	2.6	1101	17	CNS006RQ
C	43	84.4	2.6	1101	17	CNS00B01
C	44	83.8	2.6	609	17	CNS025K2
C	45	83.8	2.6	886	17	BH177277

ALIGNMENTS

RESULT 1
LOCUS AL768954
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-082A08-011867,
genomic survey sequence.
ACCESSION AL768954
VERSION AL768954.1 GI:21522073
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 Strizhov, N., Li, Y., Rosso, M., Viehoveer, P., Dekker, K., Saedler, H.
and Weisshaar, B.
A pipeline for automated high-throughput generation of FSTs
(flanking sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
Unpublished
2 Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
for flanking sequence tag based reverse genetics
Unpublished
3 (bases 1 to 473)
REFERENCE
AUTHORS Li, Y., Rosso, M., Strizhov, N. and Weisshaar, B.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer

TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Weissnar B., Max-Planck-Institut fuer
Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany

COMMENT This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone T3A5. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
Plant Genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES
source
1.473
Location/Qualifiers

/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-082A08-011867"
/note="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector pAC161. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"

BASE COUNT 168 a 58 c 70 g 146 t 31 others

ORIGIN

Query Match 4.6%; Score 146.4; DB 17; Length 473;
Best Local Similarity 90.8%; Pred. No. 1.2e-11;

Matches 167; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 1624 CTAGATGATGATTTCCAAATACATCTT-TGATGATTTAACTTAATCTTCTTC 1682
|||||
DB 186 CTGATGATGATTTCCAAATACATCTTGTGATGTTAACTCAATCTGTTCTT 127
|||||
QY 1683 CTAGGATTAATATTAATCATCGAGGTAAGTTTGTCTATTTTGGCGATGCAT 1742
|||||
DB 126 CCACTGATTAATATTAATCATCGAGGTAAGTTTGTCTATTTTGGCGATGCAT 67
|||||
QY 1743 GAAGGATTAATACCTTAATGACTTAATTTTGAAGAATGTAACCTTTTACATAGATTA 1802
|||||
DB 66 GAAGGATTAATACCTTAATGACTTAATTTTGAAGAATGTAACCTTTTACATAGATTA 7
|||||
QY 1803 TTAC 1806
|||||
DB 6 TTAC 3

RESULT 4
CNS00EVL/C 1101 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
BAC29823 of RPL-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
genoscope.

AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr

COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutyo Osogawa and
Aaron Mammeter in Piter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPL-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1.1101
Location/Qualifiers

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29823"
/clone_1lb="RPL-98"
/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match 3.5%; Score 113.2; DB 17; Length 1101;
Best Local Similarity 35.8%; Pred. No. 4e-07;
Matches 232; Conservative 118; Mismatches 297; Indels 1; Gaps 1;

QY 1625 TACATGATGATTTCCAAATACATCTTGGATGTTAACTTAATCTTCTTCCT 1684
|||||
DB 1100 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1041
|||||
QY 1685 ACCGATTAATATTAATCATCGAGGTAAGTTTGTCTATTTTGGCGATGATGA 1744
|||||
DB 1040 WATATATTAATTTTAAATAATATATATATATATATATATATATATATATAT 981
|||||
QY 1745 AGGATTAACCTTAATGATTTTGAAGAATGTAACCTTTTACTCAGATTAAT 1804
|||||
DB 980 WTAT 921
|||||
QY 1805 ACCGATGATTTTGTGCGATTAATGACAGCTGATGATGATGATGATGATGAT 1864
|||||
DB 920 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 861
|||||
QY 1865 CAATATTAATTAAGATTCATGCTACATCAATGAAAGAAACAGCTGATTAAT 1924
|||||
DB 860 TTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 801
|||||
QY 1925 TTTAATTTTAAGCAAAATTTTGAAGAAATGATATATTTTGAAGCAATATTTAATA 1984
|||||
DB 800 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 741
|||||
QY 1985 TGAATGCTATTAATGATTTTCCAT-GTTCCTAAATATTTTATTTTATTTTATTT 2043
|||||
DB 740 WTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 681
|||||
QY 2044 ATACATTAAGCAAAATAGTGTGATTCATAATATTCATTAATATTTTGAAGA 2103
|||||
DB 680 TTAATTTATTAAMWAATATTAATTAATTAATTAATTAATTAATTAATTAATTA 621
|||||
QY 2104 TCTAACAATATTAATTTAGTCAATACATGCAATGCAATGCAATGCAATGCAAT 2163
|||||
DB 620 AAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 561
|||||
QY 2164 TTAACAGAACTCCAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2223
|||||
DB 560 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 501
|||||
QY 2224 TTTTGTGGAATGAGATGATTAATATTTTATTTTATTTTATTTTATTTTATTT 2271
|||||
DB 500 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 453

RESULT 5
CNS00EVL

LOCUS	CNS000EVL	1101 bp	DNA	linear	GSS 04-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR2B23 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL069706	1	GI:4949849		
VERSION	GSS.				
KEYWORDS	Drosophila melanogaster.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a				

Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RCFI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and Est libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers

FEATURES	source	location/Qualifiers
	1. .1101	
	/organism="Drosophila	melanogaster"
	/db_xref="taxon:7227"	
	/clone="BACR29B23"	
	/clone_1lb="RRC1-98"	
	/note="end : 77"	
BASE COUNT	419 a	299 t
ORIGIN	91 c	60 g
		232 others

Query Match	3.5%	Score 110.8;	DB 17;	Length 1101;
-------------	------	--------------	--------	--------------

Matches 223; Conservative 128; Mismatches 290; Indels 5; Gaps 1;

1513 TAACCATCTGCAACTATTAATTCCTTTCTTAATGCATCTTAACATAATTATTGTTAG 1572

DD 461 TWMMWWAAIWIWWAAAWIJAIAWAAIWWAAAWAWWWATTTTWWWWWWTNA 520

[illegible][illegible][illegible]

.....

641 ||| : ||| : :: | | | | | |
700 ||| : ||| : :: | | | | | |

1753 1812

Db 701 AAAAATWAAWAAATWATAAATTTAAAAWAATAAAAWAAWAAATWAAWATAATATWAA 760

1813 TTTTGTGCCATAATGACAGCCTCTACAACTGTGATAGTCAATTTTCTGCAAAATATT 1872

Db 761 TATATATTTTAAWWAAATWWAAWWTTATAWA-----TAAATAAWAAWAAWATAAATAWATA 815

QY 1873 AATTAGGATTCATGCTACTATCAATAGAGAACAGCTGAGTATTACATTTTAATTT 19322

1873 AAATTAGGAATTCATGCTACTATCATAGAGAACAACGCTGAGTATTACATTTTAATTT 1932

Db	816	NAATMAAAMMAWAATAMNATWATATAMNATAAMAAAAAAATTAATATTAATATWATATAA	875
Qy	1933	AAAGCAAAATTTTGAAAAATGTAATAATTCTAACAAATATTAATTAATATGATGCTT	1992
Db	876	AAAAATAMATTTWTTTWTTTTMMAMWATAPAAAMATAMAAAAAAATTTTTAAAAA	935
Qy	1993	ATAATGTATTTCCATGTTCTTAATAATATTTTATATATAGTATATTAATACATTAT	2052
Db	936	WWWATATATTTTATTAATAAMTWATATWATATWATATWATATWATATWATATATATATW	995
Qy	2053	GAACCAATATATGTTGGCAATTCAAATATCTCCATTAATATTTTGAATCTCAAAAT	2112
Db	996	WTAMTAMATATATTTATTAAMWTATATTTTAAAMATATATATATATAMWTAMATATAM	1055
Qy	2113	TATTAATATTTAGTCATACATGCAATGATGCAAGATTCACAAAAA	2158
Db	1056	AAWATTAATATTAATATATATTAAMATAMAAAAAAATMAATWMAA	1101

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
AL061936	1	GI:4940214	GSS.	<i>Drosophila melanogaster</i> .	fly), genomic survey sequence.
AL061936.1	1	GI:4940214	GSS.	<i>Drosophila melanogaster</i> .	
				<i>Drosophila melanogaster</i>	
				Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
				1 (bases 1 to 1101)	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sege@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila

please see <http://www.fruitfly.org> The BDGP Drosophila

Aaron Mammoser in Pieter de Jong's laboratory in the Department of

NY. The library is named RPCI-98 and was constructed by partial reconstruction of the existing building provided by the RPCI-98.

isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library:

filters for hybridization from the BACPAC Resource Center can be used to order individual BAC clones, the entire library, or and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be

FOUND at http://backpac.med.durham.edu/aerosophila_bac.htm.

Location/Qualifiers

FEATURES

```

SOURCE
1: 1401
/organism="Drosophila melanogaster"

```

```

/clone="BACR05N11"

```

```

/next="end : TET3"

```

ORIGIN

Query Match	3.48;	Score 109.6;	DB 17;	Length 1101;
Best Local Similarity	43.98;	Pred NC 1.30-06;		

Matches	236;	Conservative	65;	Mismatches	230;	Indels	6;	Gaps	2;
---------	------	--------------	-----	------------	------	--------	----	------	----

1/48 ATAAACCTAATGACCTTAAATTTTGAAGAAATGTAACCCCTTTTACCTCAATAGATTAATTTACC 1807

[illegible][illegible]

Db 1037 TTTTMTTMTTMMMAAATTTTMCACCTTTTCATCTTCWATTTTCMTHTTTTT 978
 1865 CAATATTAATTAAGATTAATCAATGCTACTATCAATGAAGAAGACAGTATTAACAT 1924
 Db 977 AATATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 918
 1925 TTTAATTAATTAAGCAAAATTTTGAAGAAATGTTAATTAATTAATTAATTAATTAAT 1984
 Db 917 TTTAATTAATTAAGCAAAATTTTGAAGAAATGTTAATTAATTAATTAATTAATTAAT 858
 1985 TGAATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2044
 Db 857 AAAAAATTTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTT 798
 2045 TACATTAATTAAGCAAAATTTTGAAGAAATGTTAATTAATTAATTAATTAATTAAT 2104
 Db 797 AAAAAATTTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTT 741
 2105 TTTAATTAATTAAGCAAAATTTTGAAGAAATGTTAATTAATTAATTAATTAATTAAT 2164
 Db 740 TTTAATTAATTAAGCAAAATTTTGAAGAAATGTTAATTAATTAATTAATTAATTAAT 681
 2165 TTTAATTAATTAAGCAAAATTTTGAAGAAATGTTAATTAATTAATTAATTAATTAAT 2224
 Db 680 TTTAATTAATTAAGCAAAATTTTGAAGAAATGTTAATTAATTAATTAATTAATTAAT 621
 2225 TTTTGTGTAATTAAGCAAAATTTTGAAGAAATGTTAATTAATTAATTAATTAATTAAT 2281
 Db 620 TTTTGTGTAATTAAGCAAAATTTTGAAGAAATGTTAATTAATTAATTAATTAATTAAT 564

RESULT 7
 CNS01100/c 836 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
 DEFINITION BAC05N18 of Drosophila library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL099642
 VERSION AL099642.1 GI:5611253
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 836)
 Genoscope.
 Direct Submission
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billard at CERN (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelobAC11.

COMMENT
 FEATURES
 source Location/Qualifiers
 1. 836
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BAC05N18"
 /clone.lib="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : T"
 BASE COUNT 379 a 58 c 42 g 269 t 88 others
 ORIGIN

Query Match 3.1%; Score 98.4; DB 17; Length 836;
 Best Local Similarity 45.3%; Pred. No. 5.8e-05;

Matches 264; Conservative 24; Mismatches 290; Indels 5; Gaps 1;
 QY 1546 ATGCATCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1605
 Db 592 ATCCCTNNNNANNN 533
 QY 1606 CAATATTAATTAAGTACTAGTATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1665
 Db 532 TTTAATTAATTAAGTACTAGTATGATTAATTAATTAATTAATTAATTAATTAATTAAT 473
 QY 1666 CTAACTCTGTTCTTCCATAGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1725
 Db 472 TTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 413
 QY 1726 T-----TATTTTCGCATGATGAAGATTAATTAATTAATTAATTAATTAATTAATTAAT 1780
 Db 412 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 353
 QY 1781 AACCTTTTACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1840
 Db 352 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 293
 QY 1841 AACTGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1900
 Db 292 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 233
 QY 1901 AGAAGAAACAGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1960
 Db 232 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 173
 QY 1961 ATTTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2020
 Db 172 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 113
 QY 2021 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2080
 Db 112 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 53
 QY 2081 ATCTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2123
 Db 52 ATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10

RESULT 8
 CNS00E07 1101 bp DNA linear GSS 04-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
 DEFINITION BACR29P01 of RPI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL069440
 VERSION AL069440.1 GI:4949583
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see http://www.fruitfly.org The BDGP Drosophila
 melanogaster BAC library was prepared by Kazuhiro Osogawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the

COMMENT
 FEATURES
 TITLE
 JOURNAL
 ORIGIN

QY	1820	TCCCAATAAAGACAGCCTCACAACAGTGTAATGCAATTTTTCTGCAAAATATTAATG	1879
Db	612	TTTTTTTTTTTTTTTTTTTTTTTTTTTATNTTTTATTTTAAATTTTATTTTATTTM	671
QY	1880	GAATTCAAATGCTACATCAATAGAGAAACAGCTGATATCAATTTTAATTTAAAGACA	1939
Db	672	NAAATATTAATAAAAAAAAAAAAAAAAAATATATAAAAAAAAAATTTAAATTTAATAATA	731
QY	1940	AAATTTTGAAGAAAGTTATTAATTTCTACAAATATTTAAAAATATGATGCTATATGCT	1999
Db	732	TATTTTATTAATAAATATATTTAAAAAATAATATTTTATTTATTTAAAAATATATATTT	791
QY	2000	ATTTCTATGCTCTTAATAATATTTTTTTTTTTTATTTATTTAGTATTAATAATACATTAAGACCA	2059
Db	792	ATATATNNMMAATATTTTATTTTAAATTTATTTATTTATTTATTTATTTATTTATTTAATA	851
QY	2060	TAAATAGTTGGTGAATTCAAATATCTCCATTAATATTTTGAATCTACAAATTTATTAAT	2119
Db	852	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	911
QY	2120	ATTATGCAATACAAATGATGAAAGTTCCAAAAAATTTTGTTTACGAAACTTCCA	2179
Db	912	AAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	971
QY	2180	AAATTTTTTTTTTTATGCAAGCAAGAAATATACAGATAGAGAACTATTTTGTGTGGAATG	2239
Db	972	TAT	1031
QY	2240	AAGTAGTAATATACATTAAGCAAAATTTTAAAAAATATATATTAAGCCTATACGGCTCAAG	2299
Db	1032	AAAAATTTTAT	1091
QY	2300	TATGTATCT 2309	
Db	1092	CMATATTTTT 1101	
RESULT 12			
LOCUS	CNS0167M/c		
DEFINITION	CNS0167M 1201 bp DNA linear GSS 26-JUL-1999		
ACCESSION	Drosophila melanogaster genome survey sequence T7 end of BAC		
KEYWORDS	BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
VERSION	AL106396		
SOURCE	AL106396.1 GI:5621701		
ORGANISM	GSS.		
REFERENCE	Drosophila melanogaster.		
AUTHORS	Drosophila melanogaster.		
TITLE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
JOURNAL	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
	Ephydroidea; Drosophilidae; Drosophila.		
	1 (bases 1 to 1201)		
COMMENT	Genoscope.		
	Direct Submission		
	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :		
	BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
	- Web : www.genoscope.cns.fr)		
	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the European Drosophila Genome Project (EDGP) -		
	http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC		
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre		
	d'Etude du Polymorphisme Humain) with funding provided by a MRC		
	project grant. The DNA was prepared from embryos by Alain Bucheton		
	and Genevieve Payan. It has been constructed in the vector		
	pBel0BAC11.		
FEATURES			
Source	Location/Qualifiers		
	1..1201		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone="BACN15M24"		
	/clone_1lb="DrosBAC"		
	/plasmid="pBel0BAC11"		

```

BASE COUNT      323 a      87 c      79 g      551 t      161 others
ORIGIN

Query Match      2.9% ; Score 94 ; DB 17 ; Length 1201 ;
Best Local Similarity 38.4% ; Pred. No. 0.0002 ;
Matches 247 ; Conservative 87 ; Mismatches 303 ; Indels 6 ; Gaps 2

QY 1748 ATAAACCTATACCTTAAATTTTGGAAAATGTAACCCCTTTAGTCATAGATTAATTACC 1807
      |||: ||: ||: ||: ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 1166 ATAAATAATATATTANNAANNAATATAAAAAAATAATMAATATAAAMAAAMAAAMWAT 1107
      |||: ||: ||: ||: ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 1808 GATATGTTTGGTCATATATGACGCGCTCTCAAACTGTGATGACATATTTTCGCAA 1867
      ||: ||: ||: ||: ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 1106 ATTAATATMAATATATANMAAAAAAATAAAAAAAMWTTTHTANATATTTTWTWNTAT 1047
      ||: ||: ||: ||: ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 1868 ATATTAAATTTAGAAATTCATGCTACTATCAATAGACAAACAGCTGATATTACATTTT 1927
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 1046 AMAATWTTTTTTTTTTTTTTTTATAA-WMAAAAMAAAAAATTTTAAATAAATAATAT 988
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 1928 AATTTAAAGACAAAATTTTGGAAAAATGTTATTAATTTCTAACAAATATATTTAAATATGA 1987
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 987 TATMAAAATTTTAAAAATTTTWTATWTTTWTAAAAAATAAAATAATWMAAATTTT 928
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 1988 TGCCATATATGATTTTCCATGCTCTTAAAAATTTTATATTTAGTATATAATAC 2047
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 927 TTTATWTTATMAAMAAWMTTTTTTTTTTWTAAAAAATAAAATTTAAATAATATATAA 868
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 2048 ATTATGAACCAATATAGTGTGATGCAATTCATATCTCATATATATTTTGAATCTA 2107
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 867 ATAAAAAAMAAAAAAMAAAAAAMAAAAAAMWTTTATATATATAATTTTAAAAAATAA 808
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 2108 CAAATTTATTAATTTAGTCAATACAAATGCATAGAAAGTTCAAAAAATTTTGTATA 2167
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 807 AAATTTAATTTTNTTNNAAAAATAAAAAAMAAAMAAAMAAAAAATWTTAAATATAMTTAA 748
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 2168 CAGAACTCTCCAAATTTTTTTTTTTATAGCAACAAGAAATATACACATAGAAACATATTTT 2227
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 747 ATNCATTTAAACCAAAAAAAMAAATWHTATTAATTAATTAATTAATTAATTAATTT 688
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 2228 GTTGTGCAATGGAAGTAGTAATATATCATTTAAGCAATTTTAAAAA-----TTATATAG 2282
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 687 TTTTHCAAMAMAMAMAMMYMMKMCAAAAAAMMAAMMAAAMCAAMAMMTTAAAGCTAAAAA 628
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 2283 CCTATACGCGCTCAAAAGATGTATCTAGTAGTGTAATTAATATGATGCTGGCATTC 2342
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 627 CAAAMAAAAAMMAAATAAAMAAAAATCTAKKKDKKMAADAKDNKAKSAMGBCKKMAATW 568
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 2343 AGAATTTGGACACACATGAAAAACGAATTTAAATTTTAACITT 2385
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
DB 567 ACACGTGGCGCABAMAMWTTGMAITGGCAAMSAAMARRRAADDT 525
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

RESULT 13
CNS003BD/c      1101 bp      DNA      linear      GSS 03-JUN-1999
LOCUS
DEFINITION      Drosophila melanogaster genome survey sequence TEI3 end of BAC #
                  BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit
                  fly), genomic survey sequence.
ACCESSION      AL064091
VERSION      AL064091.1 GI:4941847
SOURCE      GSS.
ORGANISM      Drosophila melanogaster.
                  Drosophila melanogaster.
                  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                  Ephyridioidea; Drosophilidae; Drosophila.
                  1 (bases 1 to 1101)
                  Genoscope.
                  Direct Submission
REFERENCE      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
                  BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                  - web : www.genoscope.cns.fr)
JOURNAL

```


COMMENT

determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the p1 and ESR libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR08K08"
/clone.lib="RPCI-98"
/note="end : TET3"
BASE COUNT 395 a 120 c 103 g 334 t 149 others
ORIGIN

Query Match 2.9%; Score 93.6; DB 17; Length 1101;
Best Local Similarity 42.8%; Pred. No. 0.00024;
Matches 205; Conservative 51; Mismatches 218; Indels 5; Gaps 1;

```
QY 1722 TGTCTTATTTGCGATGCGATGAGATTAACCTATGACTTATTTGAAATGTA 1761
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1062 TATATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1003
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1782 ACCCTTTACGATGATTAATACCGATGTTTGTGCGATTAAGACAGCCTGACA 1841
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1002 ATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 943
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1842 ACTGTGATGATGATTTTCTGCAATATTAATTTAGGATTCGATCTCATATA 1901
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 942 AAATAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 883
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1902 GAGAAACAGCTGATATTTATTTATTTATTTATTTATTTATTTATTTATTT 1956
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 882 WAAATATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 823
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1957 TATAATTTCTAACAATTTATTTATTTATTTATTTATTTATTTATTTATTT 2016
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 822 TATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 763
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 2017 AATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2076
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 762 WTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 703
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 2077 AATATTTCTCATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 2136
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 702 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 643
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 2137 GCATAGAAAGTCCAAAAAATTTGTTGCAAGAACTCCAAATTTTATTTAT 2195
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 642 AATATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 584
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

RESULT 14

CNS00E07 1101 bp DNA linear GSS 04-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence TET3 and of BAC: BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL069440.1 GI:4949583

VERSION GSS

KEYWORDS Drosophila melanogaster.

SOURCE Drosophila melanogaster.

REFERENCE

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the p1 and ESR libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29P01"
/clone.lib="RPCI-98"
/note="end : TET3"
BASE COUNT 366 a 66 c 104 g 351 t 214 others
ORIGIN

Query Match 2.9%; Score 93.6; DB 17; Length 1101;
Best Local Similarity 40.4%; Pred. No. 0.00024;
Matches 246; Conservative 83; Mismatches 269; Indels 11; Gaps 3;

```
QY 1866 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1925
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 425 ACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 484
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1926 TTAATTTAAGACAAATTTTGAATATTTATTTATTTATTTATTTATTTAT 1965
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 485 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 544
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1986 GATGCTATTAATTAATTTCC-TATGTTCTTAATTAATTTTATTTATTTAT 2044
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 545 TATTAATTAATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTT 604
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 2045 TACATTTATGACCAATTAATTTATTTATTTATTTATTTATTTATTTATTT 2104
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 605 WAAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 664
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 2105 CTACAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2164
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 665 AATATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 719
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 2165 TAACAGAACTTCCAAATTTTATTTATTTATTTATTTATTTATTTATTTAT 2219
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 720 TAAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 779
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 2220 ACTATTTTGTGCGATGAGTATTAATTAATTAATTAATTAATTAATTAAT 2279
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 780 WAAAAAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 839
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 2280 AAGCCTATGCGCTCAAGATGTTATTTATTTATTTATTTATTTATTTATTT 2339
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 840 TWTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 899
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 2340 TTCAATTTGGGACACAAATGAAATTAATTAATTAATTAATTAATTAAT 2399
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 900 AATGTTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 959
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

QY	2400	TTTGGCTAAATGTCTTTTCGACTTTGAGGGGCAAAAAAAAAAGACAATCCCAAAAGTCA	2459
Db	960	TATWAAATATWTGTATTTAAATATWGRATTTTAAAAAATAATATWAAATATKKT	1019
QY	2460	CGGGTTTGA	2468
Db	1020	GTTTWTKMA	1028
RESULT 15			
LOCUS	CNS00B01/c	1101 bp	DNA
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR24D09 of RPci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL057419		
VERSION	AL057419.1	GI:4937885	
KEYWORDS	GSS.		
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)		
AUTHORS	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The_BDGP_drosophila_melanogaster_BAC_library		
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage		
JOURNAL	BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)		
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The_BDGP_drosophila_melanogaster_BAC_library		
FEATURES			
source	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR24D09" /clone_1lb="RPci-98" /note="end : 77"		
BASE COUNT	420 a 56 c 48 g 261 t 316 others		
ORIGIN			
Query Match	2.9% Score 93; DB 17; Length 1101;		
Best Local Similarity	39.2% Pred. No. 0.00029;		
Matches 235; Conservative 78; Mismatches 278; Indels 8; Gaps 1;			
QY	1809	TATGTTTTGGCCATPANTGCACCCCTCAACTGATGATGCAATTTTTTCGCAA	1868
Db	1096	TTTMTTTTMAATATWAAAATATTTATATTTTATTTATTTTTTTTTTTWTATWTA	1037
QY	1869	TATTAATTTAGCAATTCATGCTACTATCAATAGAAAGACAGCTATTTACATTTTA	1928
Db	1036	TWATTAATATWATTTTAAAAAATWAAATWTTTAAATTAATWAAATTTTAAATWTTT	977
QY	1929	ATTTAAGCAAAATTTTGAATAATGTTATATTTCTTACACATATTTTAAATATGAT	1988
Db	976	TWATTTTWTWATTTTAAATTTATTTATTTATTTATTTAAATTAATWAAATWAAATTAAT	917
QY	1989	GCCCTATATGATTTCCATGTCCTATAAATATTTTATATATTTAGTATATAATACA	2048

[illegible]

```
Search completed: March 30, 2003, 08:53:30
Job time : 4211.19 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:06:10 ; Search time 3525.39 Seconds

(without alignments)
2955.362 Million cell updates/sec

Title: US-09-502-426a-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagt.....tgcataatcacaacacaag 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_com:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	358	100.0	4818	8 AF044216	AF044216 Arabidops
2	356.4	99.6	84196	8 AT73A5	AL132979 Arabidops
3	121	33.8	1691	8 AF412114	AF412114 Arabidops
4	58	16.2	5449	6 AX346543	AX346543 Sequence
5	58	16.2	7560	6 AX346125	AX346125 Sequence
6	57.2	16.0	173786	2 AC107582	AC107582 Rattus no
7	56.6	15.8	309233	2 AC098557	AC098557 Rattus no
8	56.4	15.8	5926	6 AX346554	AX346554 Sequence
9	55.6	15.5	12356	6 AX251264	AX251264 Sequence
10	55.6	15.5	169546	2 AC004157	AC004157 Plasmodiu
11	55	15.4	7351	6 AX344930	AX344930 Sequence
12	55	15.4	20486	6 AX281500	AX281500 Sequence
13	54.8	15.3	349980	6 AC116967	AC116967 Dictyoste
14	54.8	15.3	349980	6 AC144570	AC144570 Sequence
15	54.6	15.3	123280	2 AC117076	AC117076 Dictyoste
16	54.6	15.3	268147	2 AC116966	AC116966 Dictyoste
17	54.2	15.1	2814	6 AX347029	AX347029 Sequence
18	54.2	15.1	6657	3 AC114263	AC114263 Dictyoste
19	54.2	15.1	67919	2 AC123610	AC123610 Mns muscu
20	54.2	15.1	204652	2 PFMAL13P6	AL049183 Plasmodiu
21	54	15.1	7138	6 AX333841	AX333841 Sequence
22	54	15.1	187013	2 AC116920	AC116920 Dictyoste
23	53.2	14.9	6047	2 AC115582	AC115582 Dictyoste
24	53.2	14.9	14147	6 AX251501	AX251501 Sequence
25	53.2	14.9	14147	6 AX347392	AX347392 Sequence
26	53.2	14.9	14147	6 AX349113	AX349113 Sequence
27	53	14.8	7261	6 AX251434	AX251434 Sequence
28	53	14.8	133501	2 AC116956	AC116956 Dictyoste
29	53	14.8	349980	6 AX344567	AX344567 Sequence
30	52.8	14.7	5678	6 AX346040	AX346040 Sequence
31	52.8	14.7	107739	2 AC116979	AC116979 Dictyoste
32	52.8	14.7	256172	2 AC005139	AC005139 Plasmodiu
33	52.8	14.7	310779	2 AC005140	AC005140 Plasmodiu
34	52.6	14.7	6657	3 AC114263	AC114263 Dictyoste
35	52.6	14.7	24091	2 AC014610	AC014610 Drosoph11
36	52.6	14.7	106993	3 AE002751	AE002751 Drosoph11
37	52.6	14.7	349980	6 AX344564	AX344564 Sequence
38	52.4	14.6	150903	2 AC113218	AC113218 Rattus no
39	52.2	14.6	6418	6 AX251768	AX251768 Sequence
40	52.2	14.6	6418	6 AX345224	AX345224 Sequence
41	52.2	14.6	16750	6 AX251068	AX251068 Sequence
42	52.2	14.6	16750	6 AX345423	AX345423 Sequence
43	52.2	14.6	66441	3 PFMAL1P4	AL031747 Plasmodiu
44	52.2	14.6	234112	3 PFMAL1P2	AL035475 Plasmodiu
45	51.8	14.5	627	11 CNS06EMH	AL395655 T7 end of

ALIGNMENTS

RESULT 1
AF044216 4818 bp DNA linear PLN 25-JUN-2001
LOCUS Arabidopsis thaliana steroid 22-alpha-hydroxylase (DWFA) gene,
DEFINITION complete cds.
ACCESSION AF044216
VERSION AF044216.1 GI:2935341
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 4818)
Choe,S., Dillkes,B.P., Fujioka,S., Takatsuto,S., Sakurai,A. and

Pred. No. is the number of results predicted by chance to have a

TITLE The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates multiple 2alpha-hydroxylation steps in brassinosteroid biosynthesis
 JOURNAL Plant Cell 10 (2), 231-243 (1998)
 MEDLINE 98158690
 PUBMED 9490746
 REFERENCE 2 (bases 1 to 4818)
 AUTHORS Choe, S., Dikes, B.P., Azpilroz, R. and Feldmann, K.A.
 TITLE Direct Submission
 JOURNAL Submitted (22-JAN-1998) Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
 FEATURES
 source Location/Qualifiers
 1. 4818
 /organism="Arabidopsis thaliana"
 /strain="Ws-2"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="8 cm from nga6"
 990. 4398
 /gene="DWF4"
 /note="dwarf4"
 /allele="dwarf4-1; dwarf4-2; dwarf4-3; dwarf4-4"
 990. 1055
 /gene="DWF4"
 join(1133..1353,1434..1758,1844..1996,2095..2409,2562..2654,2746..2824,2931..3040,3795..4398)
 /gene="DWF4"
 /product="steroid 22-alpha-hydroxylase"
 join(1133..1353,1434..1758,1844..1996,2095..2409,2562..2654,2746..2824,2931..3040,3795..4040)
 /gene="DWF4"
 /function="brassinosteroid biosynthetic enzyme"
 /note="member of the cytochrome P450 superfamily; CYP90B1"
 /codon_start=1
 /product="steroid 22-alpha-hydroxylase"
 /protein_id="AAC05093.1"
 /db_xref="GI:2935342"
 /translation="MFEETHTLLPLLLPSLLSLLELLILKRRNRKRLNPPGKS
 GMPLEGTIGYLTATTTGDFMOQHSKYGKTYRSLFGEPTIVSADAGINRILQ
 NEGRLEFESYPRISIGILGKWSMLVLDMDHRSISLNFSLHRLRLLKDYERH
 TLFLVDSMOQNSIFSAODEAKKFTFNMAKHMDSPEETBOLKKRYVPMGCVS
 ALPLPGTAAYHKAISOAPATILKFTERKMERKIDKEDOREEYKTFDEAMKMSDH
 VRKORTDDLLGWLKHSNSTEQLDITLSLRGHETSVALMALFFLQACKAV
 EELREHLEIARAKLEGESELNMDYKMDFTCCVINEFLGNGVRLRKAKDY
 RYKGYDIPSGKWLPIVSAHLNDSRYDQPLFLPMWQOQNNNGSSSGSFTWGN
 NYMFGGGPRLCAGSELAKEIMAVFIHHLVLKFWELAEDDQFAFPVDPNGDPIR
 VSRIL"
 BASE COUNT 1566 a 727 c 888 g 1637 t
 ORIGIN
 Query Match 100.0%; Score 358; DB 8; Length 4818;
 Best Local Similarity 100.0%; Pred. No. 5e-51;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAAAGATATTTTATTTCTCTTTTGTGATATTTTAATCA 60
 Db 4041 AAAAAAAAAAAGATATTTTATTTCTCTTTTGTGATATTTTAATCA 4100
 Oy 61 TTTTGGCCCAATGATATATATAATTTGGATATATATATATGATTTGCTTT 120
 Db 4101 TTTTGGCCCAATGATATATATAATTTGGATATATATATATGATTTGCTTT 4160
 Oy 121 TTAGTTGGGTTTGAGAAAAGGTTTGCATTTGCAAGTGACGATATATAGATTGG 180
 Db 4161 TTAGTTGGGTTTGAGAAAAGGTTTGCATTTGCAAGTGACGATATATAGATTGG 4220
 Oy 181 GAGCTAGGTTGAGCTTTGGACATTTGATTTGATTTGATTTAGTTCGACACT 240
 Db 4221 GAGCTAGGTTGAGCTTTGGACATTTGATTTGATTTAGTTCGACACT 4280
 Oy 241 ATTAAACCTTAATGGGCTTTCTATAGGCCCAATTATATATGATTAACAAGTGAC 300
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 4281 ATTAAACCTTAATGGGCTTTCTATAGGCCCAATTATATATGATTAACAAGTGAC 4340
 Oy 301 AACTTTACTCTGTTTGTGATCGGAACATTAACAAATTTGCAATTCACAAACAG 358
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 4341 AACTTTACTCTGTTTGTGATCGGAACATTAACAAATTTGCAATTCACAAACAG 4398
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 RESULT 2
 AT33A5
 LOCUS Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
 DEFINITION
 ACCESSION AL132979
 VERSION AL132979.2 GI:6782244
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 84196)
 Bloecker, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Queller, F. and
 Salanoubat, M.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 84196)
 AUTHORS EU Arabidopsis sequencing project.
 TITLE Direct Submission
 JOURNAL Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopfersplitz 18a, D-82152 Martinsried, FRG, E-mail:
 lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project
 Coordinator: Marcel Salanoubat and Francis Queller, Groupement
 d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
 Gaston Cremlieux, BP191, 91060 Evry Cedex, France;
 http://www.genoscope.cns.fr
 On Jan 27, 2000 this sequence version replaced gi:6434247.
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
 FEATURES
 source 1..84196
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 2983..3338
 /note="355 bp LTR"
 3339..7564
 /note="LTR-transposon"
 7565..7922
 /note="367bp LTR"
 12339..13197
 /gene="T3A5.10"
 join(12339..12665,12742..12803,12888..13055,13125..13197)
 /gene="T3A5.10"
 /note="similarity to cyclin-dependent kinase inhibitor
 (ICK1), Arabidopsis thaliana, EMBL:AT094772.1"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB62432.1"
 /db_xref="GI:6561966"
 /translation="MAVRRRRDVEENGVTYTTVKKRKKMEEDVDSRIILSPV
 QATNNGIVARRSACASETSYIVRRRSPPEEQQLIEEDSVSCSTSEKSR
 IEEVDLENNGDRETERISWYDILNKSEESNMDSVAVDVSRRLRLSLHETV
 KEALDEDFVOVAENKDLRNKLLCSMKYNFDEKDEPLGGREYENKLP"
 12339..12665
 /gene="T3A5.10"
 /number=1
 12666..12741
 /gene="T3A5.10"
 /number=1
 12742..12803
 /gene="T3A5.10"
 /number=2
 12804..12887
 Intron
 exon
 exon
 Intron
 Intron

```

/gene="T3A5.10"
/number=2
12888..13055
/gene="T3A5.10"
/number=3
13056..13124
/gene="T3A5.10"
/number=3
13125..13197
/gene="T3A5.10"
/number=4
15858..16470
/gene="T3A5.20"
complement(join(15858..15901,16014..16470))
/notes="similarity to Predicted proteins, Arabidopsis
thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB62433.1"
/db_xref="GI:6561967"
/translation="MARLEHNSMNGCARISFSENEFEIRSEKSNKSNINISRSFS
MSADPASFVSYSMTPADEITLKGKILPEKETSVMHRTIGEELLTEEGSVADGNTF
SRPILSSSFSTGTWRELLGLKTRHVRSKTDKVEEVLSDHKILISGVATREC
OVAOTR"
complement(15858..15901)
/number=1
complement(15902..16013)
/number=1
complement(16014..16470)
/number=2
18303..19931
/gene="T3A5.30"
complement(18303..19931)
/codon_start=1
/product="scarecrow-like 7 (SCL7)"
/protein_id="CAB62434.1"
/db_xref="GI:6561968"
/translation="MAYWCTDGNLMAIAOOLIKOQOQOQOHOQOHOEOPHPM
PSFETLPESGDFPQVYNDPGEHFPHEHONAAVAASEPDEMEESTLNGDAS
QTNDFPIYGHDPVFPSPRLSAPSYLNRKAKDSAOQLPPASTAIPSPSPQ
HPPPPPOPDPDLNOLPIKATHOYARKEPPTDLIRIKESVSGDPIGVSEFE
ALSHKTESPSSSSSSLEDPITLSKTLNACPSKPAHLANQALILEANQNNHII
AIIIDNEPTFVLTPIQLNGSSFRVDPDEVLVNFMLELYKLIDETATVGAIRLA
RSLNPRIVTIGEYEVSLNRYEFANRYKNSLREYSAVESELEPNLDRSKRLAVEYL
GRRIMDLVRSDDNNKPGTRFGIMEEKEDQRMVLMKAGPEPYKPSNVAVSQAKLLM
NRYNSTLSVESPPGIFSLAMNVPILLTVSSWR"
complement(18303..19931)
/number=1
26093..28999
/gene="T3A5.40"
complement(join(26093..26338,27091..27200,27308..27386,
27478..27570,27723..28037,28136..28288,28374..28698,
28779..28999))
/notes="Contains Cytochrome P450 cysteine heme-liron ligand
signature AA455-464:Prokaryotic membrane lipoprotein lipid
attachment site AA320-330"
/codon_start=1
/product="steroid 22-alpha-hydroxylase (DWF4)"
/protein_id="CAB62435.1"
/db_xref="GI:6561969"
/translation="MFEYHTHTLLPLLPISLILPILILKRRNRKTRNLPKGS
GMPFLGETIGYLKPYATLIDGPMOAHYSKYKITYRNLFGEPITVSADAGLNRIILO
NGRLEPESYPSIGIGILGKSMYLTVGDMDHMDMSISLNFSLHARLTLILKDEYERH
TLFLVDSMOONSIFSNQDPAKFTTNLAKKHIMSDPGEETEQLKKEYVFMKGVS
APILNPGTRYHKAQSRAITLFIERKMEERLIDJKEDEOEERETKDEBAMSDH
VKORITDDLLGWLKHSNLSTEOILIDJLILFAGHETSSVAIALATEFQACQKAY
EBLREHLEIARAKKEGSELMDYKMKMPDTCOVIMETRLIGVAVFPLRKALQDV
RYKGYDIPSGWVLPVIAVHLNDSRYOPLNFNRWROOONNGSSGSSFTSWGN
NYMPGGPRLCAGSELAKLEAVYFIHHLVLKFNELLEDKPAFPPVDPNGLPIR
VSRIL"
complement(26093..26338)
/number=1

```

```

intron complement(26339..27090)
/number=1
exon complement(27091..27200)
/number=2
intron complement(27201..27307)
/number=2
exon complement(27308..27386)
/number=3
intron complement(27387..27477)
/number=3
exon complement(27478..27570)
/number=4
intron complement(27571..27722)
/number=4
exon complement(27723..28037)
/number=5
intron complement(28038..28135)
/number=5
exon complement(28136..28288)
/number=6
intron complement(28289..28373)
/number=6
exon complement(28374..28698)
/number=7
intron complement(28699..28778)
/number=7
exon complement(28779..28999)
/number=8
38307..41323
/gene="T3A5.50"
complement(join(38307..38796,38877..39090,39189..39270,
40180..40242,40337..40401,40489..40564,40655..40734,
40805..40919,41225..41323))
/notes="Contains Eukaryotic putative RNA-binding region
RNP-1 signature AA179-186"
/codon_start=1
/product="U1 snRNP 70K protein"
/protein_id="CAB62436.1"
/db_xref="GI:6561970"
/translation="MDSGDPEFLRPNMAVQARAKVONRANVLOLKLKGSHPGTATN
NLKLEPPEPPEYKPPPEKRCQPYTGACQFVSNEAPGDEYAPPKPEYELDSQR
EPIHKLRIKGVKAEADLKTDPNNDPAATGDPYKTLFVRLRYESESSEKIKPEFS
VPIKRVHIVTQNLNPKPGYAFLEYMTRDKAKAYKADQKIDGRVAVDVGRT
VNMNPRIVTGGIGLGSVRVGGEGEIVGEOQPOGRTSOSSEPSPREERKSKKEKE
RSRELSHEDPREPSRDRREDKHNRDQGDGRDRDRDRDRDRDRDRDRDRDRDR
GSDRTSRDHDNRSKKEEDYEGGEYEHGGGRSREDAEYRGPEETRGYYEDQGD
TDYRSRDKMEEDDFRYEREKRSKRSREYVR"
complement(38307..38796)
/number=1
intron complement(38797..38876)
/number=1
exon complement(38877..39090)
/number=2
intron complement(39091..39188)
/number=2
exon complement(39189..39270)
/number=3
intron complement(39271..40179)
/number=3
exon complement(40180..40242)
/number=4
intron complement(40243..40336)
/number=4

```

Query Match 99.6%; Score 356.4; DB 8; Length 84196;
Best Local Similarity 99.7%; Pred. No. 4,7e-51;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 AAAAAAAAAAGAGAAAGTATTTTATCTCTCTTTTGTGTAATTTAAATCA 60
db 26092 AAAAAAAAAAGAGAAAGTATTTTATCTCTCTTTTGTGTAATTTAAATCA 26033
OY 61 TTTTTCCTCCCAATGATATATAAATTTGATTAATATATATGATATTCGTTT 120

```

DB 26032 TTTTTCCTCCCAATGATATATAAATTTGGTAAATATATATATGATATCTTTT 25973

QY 121 TTAGTTCGGGTTTGGAGAAAAGGTTTCGACTTTCGAAAGTCAGATATATAGATTGG 180

DB 25972 TAGAGTCGGGTTTGGAGAAAAGGTTTCGACTTTCGAAAGTCAGATATATAGATTGG 25913

QY 181 GAGCTAGGTTGAGCTCTTTCGACATTTGTATTTGAGATCTTTGATTTATTTAGTGTGACACT 240

DB 25912 GAGCTAGGTTGAGCTCTTTCGACATTTGTATTTGAGATCTTTGATTTATTTAGTGTGACACT 25853

QY 241 ATTAACCTTAATGGGCTTTCTATTAAGGCCCAATATATATGATATATACAAAGTGAC 300

DB 25852 ATTAACCTTAATGGGCTTTCTATTAAGGCCCAATATATATGATATATACAAAGTGAC 25793

QY 301 AACTTTTACCTGCTTTTTCGATCCGAGCAATTAACAATTTGTCAAAATCCAAACACACAG 358

DB 25792 AACTTTTACCTGCTTTTTCGATCCGAGCAATTAACAATTTGTCAAAATCCAAACACACAG 25735

RESULT 3
AF412114 1691 bp mRNA linear PLN 24-SEP-2001

LOCUS Arabidopsis thaliana AT3g50660/T3A5_40 mRNA, complete cds.

DEFINITION AF412114

ACCESSION AF412114.1 GI:15724347

VERSION

KEYWORDS FLI_CDNA.

SOURCE Arabidopsis thaliana.

ORGANISM Arabidopsis thaliana.

REFERENCE 1 (bases 1 to 1691)

AUTHORS Cheuk, R., Chen, H., Kim, C.-J., Koesema, E., Meyers, M. C., Banh, J., Bowser, L., Carninci, P., Dale, J. M., Goldsmith, A. D., Hayashizaki, Y., Ishida, J., Jiang, P. X., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J. M., Lin, J., Liu, S. X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C. S., Palm, C. J., Pham, P. K., Quach, H. L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C. C., Tortum, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R. W., Theologis, A. and Ecker, J. R.

TITLE Arabidopsis cDNA clones

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1691)

AUTHORS Cheuk, R., Chen, H., Kim, C.-J., Koesema, E., Meyers, M. C., Banh, J., Bowser, L., Carninci, P., Dale, J. M., Goldsmith, A. D., Hayashizaki, Y., Ishida, J., Jiang, P. X., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J. M., Lin, J., Liu, S. X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C. S., Palm, C. J., Pham, P. K., Quach, H. L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C. C., Tortum, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R. W., Theologis, A. and Ecker, J. R.

TITLE Direct Submission

JOURNAL RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : "RIKEN Arabidopsis Full-length cDNA") : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

COMMENT The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Cheuk, R., Chen, H., Kim, C.-J., Koesema, E., Meyers, M. C., Shinn, P., Banh, J., Bowser, L., Dale, J. M., Goldsmith, A. D., Jiang, P. X., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J. M., Lin, J., Liu, S. X., Miranda, M., Nguyen, M., Onodera, C. S., Palm, C. J., Pham, P. K., Quach, H. L., Southwick, A., Tang, C. C., Tortum, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R. W., Theologis, A., and Ecker, J. R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to

FEATURES
source this work. Shinozaki, K. (RIKEN GSC) and Ecker, J. R. (SSP/Salk) contributed equally to this work as PIs.
Location/Qualifiers
1..1691
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RAF106-76-F19(R13551)"
/note="ecotype: Columbia"
1..28
29..1570
/note="steroid 22-alpha-hydroxylase (DMF4)"
/codon_start=1
/product="AT3g50660/T3A5_40"
/protein_id="AA106567.1"
/db_xref="GI:15724348"
/translation="MPEEHHLPLLLPSLLSLPLILKRRNRKRFNLPCKS
GPRFEGRTGYKPKPTATTLGCPMOHNSKYKRTIRSNIFGEPTIYSADAGINRFLQ
NGLRLECSYPSIGIGLQKMSLVGVGMHMDMSISLNFISHARLRLIKDVERH
TLFVLDWQONSIFSAODEAKKFTFNLMKHMSPDEEEOLEKREYTEMKGVSA
APLNLPGTAHYHRALOSRAVILKFIERKMERKLDIKEEOEEVEVTEDEAKSKSDH
VKORHDDDLGVLKHSNLSTEQILDLITSLFAGHETSSVAIALAIFPLQACPKAV
BELREHEIARAKKEIGSESLNMDYKMDPTOCYINETLGLGVNVRHLRAKLDV
RKGYDIPSGWKVPLVYSVHVDNSRYQPNLPNPRMQQNNKASSSSGSGFSTGN
NMPFGGGRCLAGSELAKLENAVFIHHLVLFNMLAEDDPFAFPVDFPGLPIR
VSRIL"
1571..1691

BASE COUNT 3'UTR 524 a 286 c 363 g 518 t

ORIGIN

Query Match 33.8%; Score 121; DB 8; Length 1691;
Best Local Similarity 100.0%; Pred. No. 2.7e-11;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTGAAGATTTTATCTCTCTTTTGTGATATTTTAATCA 60
DB 1571 AAAAAAAAAAGTGAAGATTTTATCTCTCTTTTGTGATATTTTAATCA 1630

QY 61 TTTTTCCTCCCAATGATATATAAATTTGATTAATATATATGATATTCGTTT 120
DB 1631 TTTTTCCTCCCAATGATATATAAATTTGATTAATATATATGATATTCGTTT 1690

QY 121 T 121

DB 1691 T 1691

RESULT 4
AX346543 5449 bp DNA linear PAT 01-FEB-2002

LOCUS Sequence AX346543

DEFINITION AX346543

ACCESSION AX346543.1 GI:18494429

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1

AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.

TITLE Diagnosis of diseases associated with the immune system

JOURNAL Patent: WO 0200928-A 1614 03-JAN-2002;
Epigenomics AG (DE)

FEATURES
source Location/Qualifiers
1..5449
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN
BASE COUNT 1146 a 90 c 1299 g 2914 t

Query Match 16.2%; Score 58; DB 6; Length 5449;
Best Local Similarity 65.4%; Pred. No. 0.68;
Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Assembly program: Phrap, version 0.990329
Consensus quality: 85701 bases at least Q40
Consensus quality: 94901 bases at least Q30
Consensus quality: 103067 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 92 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1044: contig of 1044 bp in length
1045 1144: gap of unknown length
1145 2264: contig of 1120 bp in length
2265 2364: gap of unknown length
2365 3650: contig of 1286 bp in length
3651 3750: gap of unknown length
3751 4927: contig of 1177 bp in length
4928 5027: gap of unknown length
5028 6190: contig of 1163 bp in length
6191 6290: gap of unknown length
6291 7306: contig of 1016 bp in length
7307 7406: gap of unknown length
7407 8489: contig of 1083 bp in length
8490 8590: gap of unknown length
8590 9779: contig of 1189 bp in length
9779 9878: gap of unknown length
9878 11183: contig of 1305 bp in length
11183 11283: gap of unknown length
11283 12388: contig of 1105 bp in length
12388 12488: gap of unknown length
12488 13543: contig of 1055 bp in length
13543 13643: gap of unknown length
13643 14803: contig of 1160 bp in length
14803 14903: gap of unknown length
14903 16009: contig of 1106 bp in length
16009 16109: gap of unknown length
16109 17310: contig of 1201 bp in length
17310 17410: gap of unknown length
17410 18416: contig of 1006 bp in length
18416 18516: gap of unknown length
18516 19745: contig of 1229 bp in length
19745 19845: gap of unknown length
19845 20945: contig of 1100 bp in length
20945 21045: gap of unknown length
21045 22132: contig of 1087 bp in length
22132 22232: gap of unknown length
22232 23496: contig of 1264 bp in length
23496 23596: gap of unknown length
23596 24728: contig of 1132 bp in length
24728 24828: gap of unknown length
24828 25982: contig of 1154 bp in length
25982 26082: gap of unknown length
26082 27212: contig of 1130 bp in length
27212 27312: gap of unknown length
27312 28536: contig of 1224 bp in length
28536 28636: gap of unknown length
28636 29865: contig of 1229 bp in length
29865 29965: gap of unknown length
29965 31115: contig of 1150 bp in length
31115 31215: gap of unknown length
31215 32311: contig of 1096 bp in length
32311 32411: gap of unknown length
32411 33460: contig of 1049 bp in length
33460 33560: gap of unknown length
33560 34759: contig of 1199 bp in length
34759 34859: gap of unknown length
34859 35937: contig of 1078 bp in length
35937 36037: gap of unknown length

36038 37158: contig of 1121 bp in length
37158 37258: gap of unknown length
37258 38454: contig of 1196 bp in length
38454 38554: gap of unknown length
38554 39786: contig of 1232 bp in length
39786 39887: gap of unknown length
39887 40916: contig of 1039 bp in length
40916 41015: gap of unknown length
41015 42218: contig of 1103 bp in length
42218 43243: contig of 1025 bp in length
43243 43344: gap of unknown length
43344 44733: contig of 1390 bp in length
44733 44834: gap of unknown length
44834 46747: contig of 1914 bp in length
46747 46848: gap of unknown length
46848 48885: contig of 2038 bp in length
48885 48985: gap of unknown length
48985 50103: contig of 1118 bp in length
50103 50203: gap of unknown length
50203 51246: contig of 1043 bp in length
51246 51346: gap of unknown length
51346 52475: contig of 1129 bp in length
52475 52575: gap of unknown length
52575 54246: contig of 1671 bp in length
54246 54346: gap of unknown length
54346 55700: contig of 1354 bp in length
55700 55800: gap of unknown length
55800 57340: contig of 1540 bp in length
57340 57441: gap of unknown length
57441 58866: contig of 1426 bp in length
58866 58966: gap of unknown length
58966 60262: contig of 1236 bp in length
60262 60362: gap of unknown length
60362 62046: contig of 1684 bp in length
62046 62146: gap of unknown length
62146 63851: contig of 1705 bp in length
63851 63951: gap of unknown length
63951 65247: contig of 1236 bp in length
65247 65347: gap of unknown length
65347 66839: contig of 1492 bp in length
66839 66939: gap of unknown length
66939 68082: contig of 1143 bp in length
68082 68182: gap of unknown length
68182 69518: contig of 1336 bp in length
69518 69618: gap of unknown length
69618 71049: contig of 1431 bp in length
71049 71149: gap of unknown length
71149 72450: contig of 1301 bp in length

Query Match 16.0%; Score 57.2; DB 2; Length 173786;
Best Local Similarity 62.2%; Pred. No. 0.4;
Matches 89; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGTGAATATTTATCTCTCTTTTTCATATTTGAATATTTAAATGA 60
DB 27835 AAAAAAAAAAAGTGAATATTTATCTCTCTTTTTCATATTTGAATATTTAAATGA 27894
QY 61 TTTTTCCTCCATGATATATTAATTTGATATTAATTTGATATTTGTTT 120
DB 27895 ATTTTTCCTCCATGATATATTAATTTGATATTAATTTGATATTTGTTT 27954
QY 121 TTAGTTCGGGTTGAGAAAAGG 143
DB 27955 TTAATTAANGGGGGGAAAAGG 27977

RESULT 7
LOCUS AC098557 309233 bp DNA linear HTG 12-JUL-2002
DEFINITION Rattus norvegicus clone CH230-81P10, *** SEQUENCING IN PROGRESS
ACCESSION AC098557

VERSION	AC098557.6	GI:21728854
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	Norway rat.	

ORGANISM	REFERENCE
<i>Rattus norvegicus</i>	1 (bases 1 to 309233)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
Muzny, D.M., Adams, C., Adio-Obiola, B., et al.	

Albrooks, S. L., Amarantunge, H. C., Are, J. R., Ayale, M., Banks, C.,
Barbaria, J., Benton, J., Blinige, K., Blankenburg, K., Bonini, D.,
Bouck, J., Bowle, S., Brileva, M., Brown, E., Brown, M., Bryant, N. P.,
Buhay, J., Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C.,
Carroll, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R.,
Davis, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A.,
Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H.,
Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J.,
Earhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, C. C., Elhaj, C., Escotto, M.,
Garrett, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homsli, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L. E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., Kling, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, J.,
Li, J., Li, E., Lichtarge, O., Lien, C., Liu, J., Liu, W., Louisaged, H.,
Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Masssey, E., Mahoney, E., McLeod, M. P., Meador, M., Mel, G., Metker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogum, M., Okunnu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peterson, L., Plickens, R., Prime, E., Pu, L. L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rotjuckan, I., Rolfe, M., Ruiz, S., Seavey, G.,
Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,
Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Szatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wlezyk, R., Wooten, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE

Direct Submission
Unpublished
2 (bases 1 to 309233)
Worley/K.C.
Direct Submission
Submitted (24-Oct-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 309233)
Worley/K.C.

TITLE Direct Submission
 JOURNAL Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 11, 2002 this sequence version replaced gi:20066030.
 COMMENT ----- Genome Center

```

Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 233229 bases at least Q40
Consensus quality: 233927 bases at least Q30
Consensus quality: 234451 bases at least Q20
-----

```

```
* NOTE: Estimated segment size may differ from sequence length
* (see http://www.hpsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

*	1	1069:	contig of 1069 bp in length
*	1070	1169:	gap of unknown length
*	1170	2523:	contig of 1354 bp in length
*	2524	2623:	gap of unknown length
*	2624	3832:	contig of 1209 bp in length
*	3833	3932:	gap of unknown length
*	3933	4935:	contig of 1023 bp in length
*	4936	5055:	gap of unknown length
*	5056	6248:	contig of 1193 bp in length
*	6249	6348:	gap of unknown length
*	6349	7523:	contig of 1175 bp in length
*	7524	7523:	gap of unknown length
*	7654	8765:	contig of 1142 bp in length
*	8766	8865:	gap of unknown length
*	8866	10088:	contig of 1224 bp in length
*	10090	10189:	gap of unknown length
*	10190	11339:	contig of 1150 bp in length
*	11340	11439:	gap of unknown length
*	11440	12630:	contig of 1191 bp in length
*	12631	12730:	gap of unknown length
*	12731	13871:	contig of 1141 bp in length
*	13872	13971:	gap of unknown length
*	13972	14977:	contig of 1006 bp in length
*	14978	15077:	gap of unknown length
*	15078	16238:	contig of 1181 bp in length
*	16239	16358:	gap of unknown length
*	16359	17441:	contig of 1063 bp in length
*	17422	17551:	gap of unknown length
*	17522	18539:	contig of 1018 bp in length
*	18540	18659:	gap of unknown length
*	18660	19856:	contig of 1217 bp in length
*	19857	19956:	gap of unknown length
*	19957	20981:	contig of 1035 bp in length
*	20982	21082:	gap of unknown length
*	21082	22354:	contig of 1273 bp in length
*	22355	22455:	gap of unknown length
*	22455	23841:	contig of 1386 bp in length
*	23841	23940:	gap of unknown length
*	23941	25442:	contig of 1502 bp in length
*	25443	25543:	gap of unknown length
*	25543	26794:	contig of 1252 bp in length
*	26795	26894:	gap of unknown length
*	26895	28026:	contig of 1132 bp in length
*	28027	28127:	gap of unknown length
*	28127	29239:	contig of 1113 bp in length
*	29240	30567:	gap of unknown length
*	30568	30667:	contig of 1228 bp in length
*	30668	31810:	gap of unknown length
*	31811	31910:	contig of 1143 bp in length
*	31911	33252:	contig of 1342 bp in length
*	33253	33352:	gap of unknown length
*	33353	34543:	contig of 1191 bp in length
*	34544	34643:	gap of unknown length
*	34644	35643:	contig of 1000 bp in length
*	35644	35743:	gap of unknown length
*	35744	36743:	contig of 1000 bp in length

	Matches	118:	Conservative	0:	Mismatches	104:	Indels	0:	Gaps	0:
Oy	11	ACATGCAAGTATTTTATTCCTCTCTTTT	TTTTTTCATATATTTAAATCATTTTTTTC	70						
Db	9051	ACATTAATAGGTAGTTATTTTCTAGTTT	TTTTTGAGGAATTTTAAATGTTTTTAACT	9110						
Oy	71	CCAAATGATATATAAAAATTTGGATAAATATATTT	TGGAAATTCGTTTTTTAGTTCGGG	130						
Db	9111	GGTGTATATGATTTTATTTTATTAATAATATATAAGGG	TTTTTTTTTTTTTTTAAATATGCT	9170						
Oy	131	TTTGGAAAGGCTTTCCACTTTCGAAGTGACAGAT	TAATATAGATTTGGAGCTAGGTT	190						
Db	9171	TATTGATTTTGTATAAGCTTTGTTTTTGGATATAGTAAT	TTTAAATGGGTAAGATGAT	9230						
Oy	191	GAGTCTTGGACATTTGATTGGAGTGTGATATAGTG	232							
Db	9231	ATTATATGTAGTTTGTATTTGATTTTTCATGATTAATG	9272							

RESULT 10	AC0004157	169546 bp	DNA	linear	HTG 12-AUG-2000
LOCUS	AC0004157				
DEFINITION	Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.				
ACCESSION	AC0004157				
VERSION	AC0004157.8	GI:9797712			
KEYWORDS	HTG; HTGS; PHASE1				
SOURCE	Plasmodium falciparum.				
ORGANISM	Plasmodium falciparum.				
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
AUTHORS	1 (bases 1 to 169546)				
TITLE	Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T., Kundli,O.B., Conway,A.B., and Davis,R.W.				
JOURNAL	Plasmodium falciparum 3D7 chromosome 12				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 169546)				
TITLE	Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.				
JOURNAL	Direct Submission				
COMMENT	Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA				
	On Aug 12, 2000 this sequence version replaced cf.9810447				

```

FEATURES
    source
        2507 169346: contig of 145880 bp in length.
            Location/Qualifiers
                1..169346
                    /organism="Plasmodium falciparum"
                    /db_xref="taxon:5833"
                    /chromosome="12"
                    /clone="PFAC293"
                    /clone="3D7"
                    /clone="3D7"
BASE COUNT
69871 a 15381 c 15705 g 66389 t 200 others
ORIGIN

```

Query Match	15.5%	Score 55.6;	DB 2;	Length 169546;
Best Local Similarity	51.6%;	Pred. No. 0.74;		
Matches 127;	Conservative 0;	Mismatches 119;	Indels 0;	Gaps 0;
QY 1	AAAAAAAAAAGCATGAAGTATTTTATCTCTCTTTTGTGATATTTAAATCA	60		
DB 136495	AAAAAAAAAAGATATAACATTTGAATCTTTGTGATTTTACCTTTTTTTTT	136554		
QY 61	TTTTTTTGCCCAATGATATATAAATAATTTGGATATAAATTTTGGATTTTCGTTT	120		

Db 136555 TTTTTCCTTTTGGCTACAGAAAAATATGATTTCAATTTGTGAATTTTCCTTGA 136614
QY 121 TTATGTCGGGTTTGAGAAAAGGTTTCGACTTTGGAAGTGGACGATGTATATGATTGG 180
Db 136615 TGAATATGGAATTAAAAAACCTATGCTCATCAATAATACAAATATACCTTCCCTTTGG 136674
QY 181 GAGCTAGGTTGAGTCTTTGGACATTTTGATTTGATGATGTTGATTAGTGTGACACT 240
Db 136675 GTAAATGAAAAAAATATATACATATTATATAGACATTTATATATTATATATATA 136734
QY 241 ATTAAA 246
Db 136735 TATATA 136740

RESULT 11				
AX344930				
LOCUS	AX344930	7351 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0200928.			PAT 01-FEB-2002
ACCESSION	AX344930			
VERSION	AX344930.1	GI:18492816		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS	Olek, A., Piepenbrock, C. and Berlin, K.			
TITLE	Diagnosis of diseases associated with the immune system			
JOURNAL	Patent: WO 0200928-A 1 03-JAN-2002;			
FEATURES	Epigenomics AG (DE)			
Source	Location/Qualifiers			
	1..7351			

BASE COUNT	ORIGIN
2230	a 42 c 1230 g 3849 t

	Query Match	15.48;	Score 55;	DB 6;	Length 7351;
	Best Local Similarity	48.08;	Pred. No. 2;		
	Matches 157;	Conservative	0;	Mismatches 170;	Indels 0;
				Gaps	0
QY	20	TATTTTATTCCTCTCTTTT	TTTGGATTAATTTAAATCATTTT	TTTGGCCCAATGA	79
Db	3765	TTTGATTTTGTTATTCGTAT	GTGTGGAAAGTTT	TTTTTATTTTGGTTTCGAGANA	3824
QY	80	TATTAATAATTTGGATAAAT	TAATTAATTTGATATTCGTTT	TAGTTGGGTTTGAGANA	139
Db	3825	ATTATATTTATTAATTTT	TAGAGTTAATTTAATTTGTTT	TTTTTAAATTTTTTANA	3884
QY	140	AGGGTTTCACATTTTGC	AAAGTGAGACATTAATATAGATTGGAGCT	TAGTCTTG	199
Db	3885	AGGTTTACGTTGGTTGG	TTATGTGTTTTTATGCTGTTT	TAGAGATTTTGAATTTTTGT	3944
QY	200	GACATTTGATTTGATAGT	GTGATTAATTAATAGTGTGACACTT	TAAACTTAAAGGCT	259
Db	3945	AATAATATGTTGGTATAT	GTATTTATTAATAGTATTTTAATTTGATTTATTTGAT		4004
QY	260	TTCTTAAGGCCCAATT	TAATTAAGCATTAACAAGTGACACACTT	TAATCTGTTTTG	319
Db	4005	TTTATATAGTAGCTAT	ATATAGTTTAAATTTAAATAGTAGTGT	TTTTTGTGTTTTGT	4064
QY	320	ATCCGACGACATTAAC	AAATGCAAT	346	
Db	4065	TTTATATAGCAATGT	TAATAATTTTTTAAGT	4091	

RESULT	12
AX281500	
LOCUS	20486 bp DNA
DEFINITION	Sequence 164 from Patent WO0177376.
ACCESSION	AX281500
VERSION	AX281500.1 GI:16608755
PAT	02-NOV-2001

KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial construct.
REFERENCE
1 Olek, A., Piepenbrock, C. and Berlin, K.
AUTHORS Diagnosis of diseases associated with metastasis
TITLE Patent: WO 0177376-A 164 18-OCT-2001;
JOURNAL Epigenomics AG (DE)
FEATURES
Location/Qualifiers
source 1..20486
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 6101 a 360 c 4446 g 9577 t 2 others
ORIGIN
Query Match 15.4%; Score 55; DB 6; Length 20486;
Best Local Similarity 49.6%; Pred. No. 1.6; Indels 2; Gaps 1;
Matches 169; Conservative 0; Mismatches 170;
QY 17 AAGTATTTTATCTCTCTTTTGTGATTTTAAATCAATTTTGGCCATG 76
DB 11836 ATGTAGATTGATTTAAATATTTAGTAATATTTTAAAGTTTAAAA 11895
QY 77 ATATATTTTATTTGATTAATATTTATGATTTCTTTTACGCGTTGAG 136
DB 11896 GTAAAGAAAAATTAAGATTAATATTAAGATTTTAAAGTTTAA 11955
QY 137 AAAAGGTTTCACCTTTCGAAAGTGAGATGATATAGATTTGGAGCTGAGTCT 196
DB 11956 TTGTTTTCGAAATTTTGGAAAGCTTTTAAATTTGTTGTTTATTTT 12015
QY 197 TTGACATTTGATTTGATTTGATTTAGTGTGACACATTTAAACCTTAATG 256
DB 12016 TTAGAGTTTAAATGTTATTTAGTTAGATGAGATTAATAATTA-7TAAATTT 12073
QY 257 GCTTCTATTAAGGCCCATTTATACGTTTATTAACAAGTCAACCTTTTCTGTTT 316
DB 12074 GATTAATAATGATTTATCGTATTAAGTTAAATTAAGATTTAAATTTTATTTT 12133
QY 317 TTGATCCGACGATTAACAAATTTGCAATACCAACAAACAA 357
DB 12134 TTTATGTTAAAAAATTAATTTATTTTGAATTTAA 12174

RESULT 13
AC116967 93491 bp DNA linear HTG 04-APR-2002
LOCUS AC116967/c
DEFINITION Dictyostelium discoideum chromosome 2 map 5401525-545014 strain
AX1,*** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC116967
VERSION AC116967.1 GI:19920066
KEYWORDS HTG; HTGS; PHASE2.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE
1 (bases 1 to 93491)
AUTHORS Gloeckner, G., Eichinger, L., Szafianski, K., Pachebat, J., Dear, P.,
Lehmann, R., Baumgart, C., Parre, G., April, J. F., Gulgo, R., Kumpf, K.,
Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
Noegel, A. A.
TITLE Sequence and Analysis of Chromosome 2 of Dictyostelium
JOURNAL Unpublished
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE
2 (bases 1 to 93491)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT CDS predictions from Genaid may contain errors. Further information
is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)

and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
source 1..93491
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="5401525-5495014"
BASE COUNT 36436 a 9904 c 10019 g 37032 t 100 others
ORIGIN
Query Match 15.3%; Score 54.8; DB 2; Length 93491;
Best Local Similarity 50.4%; Pred. No. 1.2;
Matches 134; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTGAAGTATTTATCTCTTTTGTGATTTTAAATCA 60
DB 65656 AAAAAATAAAACTCAATCAATCTTATTAATTTTAAATTTCTTTT 65597
QY 61 TTTTTCCTCCCAATGATATTAATAATTTGATTAATTTATGATTTGTTT 120
DB 65596 TCTTCTATCTATTTATTTATTTTATTTTATTTTATTTTATTTT 65537
QY 121 TTAGTTCCGCTTTGAGAAAAAGGTTTCGACTTTCGAAAGTGACGATGATAGTTG 180
DB 65536 TTTGATCTAAATTTATTTATTTGATGACAGGTAATTTTATTTTATTTT 65477
QY 181 GAGCTAGTTGAGTCTTTCGACATTTGATTTGATTTGTTTATTTAGTGTGACACT 240
DB 65476 TTTATTTTATTTTATTTTATTTTAAATTTTATTTTATTTTATTTTAAAAAATA 65417
QY 241 ATTAACCTTAATGCGCTTCTATA 266
DB 65416 AATTAATAATTTATTTGCTTTTAA 65391

RESULT 14
AX344570 349980 bp DNA linear PAT 01-FEB-2002
LOCUS AX344570
DEFINITION Sequence 21 from Patent WO0200932.
ACCESSION AX344570
VERSION AX344570.1 GI:18492456
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE
1 Olek, A., Piepenbrock, C. and Berlin, K.
AUTHORS Diagnosis of known genetic parameters within the mhc
TITLE Patent: WO 0200932-A 21 03-JAN-2002;
JOURNAL Epigenomics AG (DE)
FEATURES
Location/Qualifiers
source 1..349980
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3,673,778 <223>-sp11t as follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.349.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <223>-original length of seq 2: 3,673,778 <223>-sp11t as follows.-seq 14 0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:05:35 ; Search time 71.5078 Seconds
(without alignments)
11274.505 Million cell updates/sec

Title: US-09-502-426a-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagt.....tgcataaccacaacaa 358

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	358	100.0	6888	21	AAA59599
2	58	16.2	5449	24	ABL33641
3	58	16.2	7360	24	ABL33223
4	56.4	15.8	5926	24	ABL33652
5	55.6	15.5	12356	22	AA546510
6	55	15.4	7351	24	ABL32028
7	55	15.4	20486	24	ABL34611
8	54.2	15.1	2814	24	ABL34127
9	54	15.1	7138	24	ABK28455

10	53.2	14.9	14147	22	AA546743	Tumour suppressor
11	53.2	14.9	14147	22	ABK33955	Human DNA for stag
12	53	14.8	7261	22	AA546670	Tumour suppressor
13	52.8	14.7	5678	24	ABL33138	Human immune syste
14	52.6	14.7	21231	23	ABL19002	Drosophila melanog
15	52.2	14.6	6418	24	ABL32332	Human immune syste
16	52.2	14.6	6418	24	AA561073	Human immune syste
17	52.2	14.6	16750	22	AA546314	Human gene regulat
18	52.2	14.6	16750	22	AA546314	Tumour suppressor
19	51.8	14.5	6145	24	ABL32972	Human immune syste
20	51.6	14.4	17131	24	ABL33053	Human immune syste
21	51	14.2	13584	24	ABL32615	Human immune syste
22	50.8	14.2	5311	24	ABL33019	Human immune syste
23	50.8	14.2	8087	24	ABL32742	Human immune syste
24	50.8	14.2	13606	22	AA545457	Chemically pretrea
25	50.8	14.2	13606	22	AA545457	Tumour suppressor
26	50.8	14.2	13606	24	ABL33810	Human immune syste
27	50.8	14.2	13606	24	ABK28313	DNA transcription
28	50.6	14.1	5649	22	AA546384	Tumour suppressor
29	50.6	14.1	5649	22	ABK40008	Human chemically p
30	50.6	14.1	5649	24	ABL32849	Human immune syste
31	50.4	14.1	5163	24	ABL33248	Human immune syste
32	50.2	14.0	14615	22	AA546704	Tumour suppressor
33	49.8	13.9	6134	22	AA546553	Tumour suppressor
34	49.8	13.9	6739	22	ABL32738	Human immune syste
35	49.8	13.9	8866	22	AA545433	Chemically pretrea
36	49.8	13.9	8866	24	ABK28280	DNA transcription
37	49.8	13.9	15649	24	ABL70544	Chemically treated
38	49.8	13.9	13515	24	ABL34175	Human immune syste
39	49.6	13.9	9347	24	ABL49336	Human polynucleoti
40	49.6	13.9	11422	24	ABK39936	Human chemically p
41	49.6	13.9	11422	24	ABL32218	Human immune syste
42	49.6	13.9	12639	24	ABN60106	Human chemically m
43	49.6	13.9	13427	24	ABL33926	Human immune syste
44	49.4	13.8	6609	24	ABL70526	Chemically treated
45	49.4	13.8	6609	24	AA561123	Human gene regulat

ALIGNMENTS

RESULT 1	
AAA59599	standard; DNA: 6888 BP.
AAA59599;	
14-NOV-2000	(first entry)
DNA encoding a cytochrome P450 enzyme designated DMF4.	
DMF4: cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;	
plant phenotype; cell elongation; ss.	
Arabidopsis sp.	
Key	Location/Qualifiers
promoter	1..3203
TATA_signal	/*tag- a 3060..3125
CDS	/*tag- b 3203..6110
exon	/*tag- c /product= "DMF4"
intron	/*tag- d 3203..3423
intron	/*tag- e 3424..3503
exon	/*tag- f 3504..3828
intron	/*tag- g 3829..3913
intron	/*tag- h 3914..4000

FT exon 3914..4066
 FT /tag= h
 FT Intron 4067..4164
 FT /tag= 1
 FT exon 4165..4479
 FT /tag= 1
 FT Intron 4480..4631
 FT /tag= k
 FT exon 4632..4724
 FT /tag= 1
 FT Intron 4725..4815
 FT /tag= m
 FT exon 4816..4894
 FT /tag= n
 FT Intron 4895..5000
 FT /tag= o
 FT exon 5001..5110
 FT /tag= p
 FT Intron 5111..5864
 FT /tag= q
 FT exon 5865..6110
 FT /tag= r
 FT 3'UTR 6011..6468
 FT /tag= s
 PN WO20047715-A2.
 XX 17-AUG-2000.
 XX 11-FEB-2000; 2000MO-US03820.
 XX 11-FEB-1999; 990S-0119657.
 XX 11-FEB-1999; 990S-0119658.
 XX (ARIZ-) ARIZONA BOARD OF REGENTS.
 XX Azpiroz R, Choe S, Feldmann KA;
 PI WPI: 2000-549142/50.
 DR P-PSDB; AAB07921.
 XX
 PT New isolated dwf4 polynucleotide useful for altering the phenotype of
 PT plants, for diagnostic assays and in the production of antibodies -
 XX Claim 3; Fig 10A-G; 113pp; English.
 PS
 XX The present sequence encodes a DWF4 polypeptide. The polypeptide is a
 CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
 CC brassinosteroids. Specifically, it mediates multiple
 CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
 CC polynucleotide is used for altering the phenotype of a plant. DWF4
 CC plants display a dramatic reduction in the length of different organs,
 CC and this size reduction is attributable to a defect in cell elongation.
 CC The DWF4 polynucleotides and polypeptides can be used in diagnostic
 CC assays and to generate antibodies, which can be used to produce
 CC immunogenic compositions.
 XX
 SQ Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;
 Query Match 100.0%; Score 358; DB 21; Length 6888;
 Best Local Similarity 100.0%; Pred. No. 3.2e-55;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGATGAAAGTATTTATCTCTCTTTTGGATTAATTTAAATCA 60
 DB 6111 AAAAAAAAAAGATGAAAGTATTTATCTCTCTTTTGGATTAATTTAAATCA 6170
 QY 61 TTTTTCCTCCCAATGATATATATAATTTGGATTAATTTATGATATTCGTTT 120
 DB 6171 TTTTTCCTCCCAATGATATATATAATTTGGATTAATTTATGATATTCGTTT 6230
 QY 121 TTAGTTCGGGCTTGAGAAAGGTTTCGACTTCGAAAGTGCAGATATATAGATTGG 180
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 6231 TTAGTTCGGGCTTGAGAAAGGTTTCGACTTCGAAAGTGCAGATATATAGATTGG 6290
 QY 181 GACCTAGGTTGAGCTTTGGACATTTGTATTTGATGTTGATTATAGTGCACAT 240
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 6291 GACCTAGGTTGAGCTTTGGACATTTGTATTTGATGTTGATTATAGTGCACAT 6350
 QY 241 ATTAAACCTTAATGGGCTTCTATTAAGGCCCAATTATATAGATTATTAACAAAGTGAC 300
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 6351 ATTAAACCTTAATGGGCTTCTATTAAGGCCCAATTATATAGATTATTAACAAAGTGAC 6410
 QY 301 AACTTTACTTCGTTTGTGATCCGAAAGCATATACAAATTTGCAAAATCCAAACGAC 358
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 6411 AACTTTACTTCGTTTGTGATCCGAAAGCATATACAAATTTGCAAAATCCAAACGAC 6468
 RESULT 2
 ABL33641
 ID ABL33641 standard; DNA: 5449 BP.
 AC ABL33641;
 XX
 XX 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 1614.
 XX
 XX Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosatic; neutrotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 OS
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001MO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Plepenbrock C, Berlin K;
 DR WPI: 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX Claim 1; SEQ ID NO 1614; 32pp + Sequence Listing; German.
 PS
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SQ Sequence 5449 BP; 1146 A; 90 C; 1299 G; 2914 T; 0 other;
 Query Match 16.2%; Score 58; DB 24; Length 5449;
 Best Local Similarity 65.4%; Pred. No. 0.053;
 Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
 QY 4 AAAAAAAAAAGATGAAAGTATTTATTTCTCTCTTTTGGATTAATTTAAATCATTT 63

DB 5256 AATATAAATTAAGTATTTTGAATTTTATTTTATTTTATTAATTAAGT 5315
 QY 64 TTTTGGCCCAATGATATATTAATTTGGATTAATATATGATATCGTTT 123
 DB 5316 TGTATAAGTATTAATTAATTTTATTAATTTTATTAATTTTATTAAT 5375
 QY 124 GTTCGGGTTT 133
 DB 5376 TTTTGGGTTT 5385

RESULT 3

ABL33223
 ID ABL33223 standard; DNA: 7560 BP.

AC ABL33223;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1196.

KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI: 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation

PS Claim 1; SEQ ID NO 1196; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

SO Sequence 7560 BP; 2154 A; 47 C; 1554 G; 3805 T; 0 other;

Query Match 16.2%; Score 58; DB 24; Length 7560;
 Best Local Similarity 51.6%; Pred. NO. 0.053;
 Matches 159; Conservative 0; Mismatches 145; Indels 4; Gaps 1;

QY 2 AAAAAAAAAAGATGATATTTATCTCTTTTATTTTGAATTTTAAATCAT 61
 DB 2728 ATAAATTCAGATTAATTTATTTTATTTTATTTTATTTTATTTTATTTT 2787

QY 62 TTTTGGCCCAATGATATATTAATTTGGATTAATATATTTGATATTCGTTT 121
 DB 2788 TTTTATGCGAGATATTTTAAATTTTAAAGTAAATTTTATTTATTCGTTATTT 2847
 QY 122 TAGTGGGTTTGAAGAAAGGTTTCAGTTTCGAAGTGAAGATATATAGTATG 181
 DB 2848 TGTATGATTAATTAATTAATTTTATTTTATTTTATTTTAAATTAATTAAT 2907
 QY 182 AGCTAGGTGAGTCCTTGGACATTTGATGATG---TGTGATTAATTAATGTCGAC 237
 DB 2908 TATTAATTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTTT 2967
 QY 238 ACTATTAACCTTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAAT 297
 DB 2968 ATATTTGATTAATTAATTAATTTTATTTTAAAGAAATTTATGAAATTAATTAAGA 3027
 QY 298 GACAACTT 305
 DB 3028 GATTAATTT 3035

RESULT 4

ABL33652
 ID ABL33652 standard; DNA: 5926 BP.

AC ABL33652;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1625.

KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI: 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation

PS Claim 1; SEQ ID NO 1625; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

SO Sequence 5926 BP; 1562 A; 55 C; 1318 G; 2991 T; 0 other;

[illegible]

	CC	bisphosphite, of genes associated with tumour suppression and
	CC	oncogenes having a sequence taken from 536 (actually 533 since
	CC	numbers 408, 458 and 500 are missing from the sequence listing) sequences
	CC	(Ss) and sequences complementary to (Ss). The nucleic acid may be a
	CC	peptide nucleic acid oligomer (PNA) or at least 9 nucleotides and may
	CC	form part of a set of probes for detecting the cytosine methylation state
	CC	and/or single nucleotide polymorphisms and also to be used in an
	CC	array for analysing diseases associated with CpG dinucleotides e.g.
	CC	cancers and tumours. The probes can also be used in a method for
	CC	ascertaining genetic and/or epigenetic parameters for the diagnosis
	CC	diseases by analysing cytosine methylations. The parameters may be
	CC	compared to another set of genetic and/or epigenetic parameters, the
	CC	differences serving as basis for diagnosis and/or prognosis events which
	CC	are diagnostic sequons to patients. The present sequence is one of the
	CC	533 genomic sequences derived from tumour suppressor genes and
	CC	oncogenes. Sequences with even numbered Seq ID numbers are the
	CC	complementary sequence of the corresponding odd numbered sequence (e.g.
	CC	ID 2 and ID1, ID 536 and ID 533), except for those whose partner sequence
	CC	is missing).
	CC	Note: The sequence data for this patent did not form part
	CC	of the printed specification, but was obtained in electronic
	CC	format directly from WIPO at
	CC	ftp.wipo.int/pub/published_pct_sequences.
SQ	XX	Sequence 12356 BP; 3645 A; 68 C; 2127 G; 6516 T; 0 other;
Query Match		15.5%; Score 55.6; DB 22; Length 12356;
Best Local Similarity -		53.2%; Pred. No. 0.14;
Matches 118; Conservative	0;	Mismatches 104; Indels 0; Gaps 0
OY	11	ACATGAAGTATTCTTCCTTTTTTTTGATAATTTTAANCATTTTTTTGC 70
Dd	9051	A GAATATATGGTAGCTTTAATTTTGTGAGAATTTTAAATCCTTTTATGT 9110
OY	71	CCAATCATTAATAAAAATTTGGATTAATTAATGATATTCCTTTTACTCGGG 130
Dd	9111	G GTGTATTTGATTTATTTTATTAAAAATTAATTAAGCGTTTTTTTTTTTATTTGT 9170
OY	131	T TTGAGAAAAGGGTTTCGACTTTCGAAGTCAGCAAGTATATAGATTGGAGCTAGGTT 190
Dd	9171	T ATTTAGTATTTTGTATAGTTCTTTTTCGATTAAGTATTAATTTAATTCGTAAGATGAT 9230
OY	191	G AGCTTTGGACAATTGATTTGATGATGTTGTATATTTAGTG 232
Dd	9231	A TTATATGTAGTTTGATTTGTATTTTGTGATGATTATG 9272
RESULT 6		
ABLJ32028		
ID	ABLJ32028 standard:	DNA: 7351 BP.
AC	ABLJ32028;	
XX	26-MAR-2002	(first entry)
DT		
XX	Human immune system associated gene SEQ ID NO: 1.	
DE		
KM	Human; immune system disease; cytosine methylation; antiasthmatic;	
KW	antiartherosclerotic; antianaemic; cytostatic; nootropic;	
RN	neuroprotective; anti-HIV; anticoinvulsant; ophthalmological;	
KV	antiinflammatory; antiarthritic; antididiabetic; antipsoriatic;	
KX	antinflammatorcy; cancer; eye disease; arteriosclerosis; anaemia;	
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;	
KM	neutrofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;	
gene; ds.		
OS	Homo sapiens.	
XX	MOZ00200928-A2.	
PB	03-JAN-2002.	
PD		
XX		

QY 123 AGTCCGGCTTTGAGAAAGGCTTTCACATTCGAAAGTGACATGATATGATGGGA 182
 DB 3283 AATTAGCTTTTAAAGATGATTTTGTATATATAGATTTGTTTATTTATGATGTA 3442
 QY 183 GCTAGGCTTGCTTTGACATTTGATTTGATGCTGTTGATTTATGATGACACTAT 242
 DB 3343 AATATTTTATTTTATATATTTGATTTGATTTATATTTATTTGCGGCTAA 3402
 QY 243 TAAACCTTAATGCGCTTCTATAGAGCCCAATATATATATATATATATATATAT 302
 DB 3403 TTAAGATATATTTTATATATTTGATTTGATTTGATTTGATTTTATTTATGAA 3462
 QY 303 CTTTACTTCGTTTGA 320
 DB 3463 ATTTTATAGTTTATTTAA 3480

RESULT 10
 AAS46743
 ID AAS46743 standard; DNA; 14147 BP.
 AC AAS46743;
 XX
 DT 18-DEC-2001 (first entry)
 DE Tumour suppressor gene derived chemically modified sequence #467.
 XX
 KW Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; Cpg dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PE 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 OS
 PS Claim 1: SEQ ID No 467; 27pp: English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with Cpg dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be

CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp://ipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 14147 BP; 3621 A; 225 C; 3116 G; 7185 T; 0 other;
 Query Match 14.9%; Score 53.2; DB 22; Length 14147;
 Best Local Similarity 48.1%; Pred. No. 0.36;
 Matches 151; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
 QY 7 AAAAAAGTGAAGATTTTATCTCTCTTTTATTTTATTTTAAATCATTTT 66
 DB 10811 AAAAAAGTGTGTTAGTTGTTGCTTTGTTTATTTTATTTTATTTTATTTT 10870
 QY 67 TTGCCCAATGAT 126
 DB 10871 TTTTTCGAATTTATATGTTTATTTTATTTTATTTTATTTTATTTTATTTT 10930
 QY 127 CGGCTTGAGAAAAGGTTTCGACTTTCGAAAGTGACAGATATATATATATAT 186
 DB 10931 TTGGGATTTTAT 10990
 QY 187 GGTGAGTCTTTGACATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 246
 DB 10991 TTTAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 11050
 QY 247 CCTTAATGCGCTTCTATAGAGCCCAATATATATATATATATATATATATATAT 306
 DB 11051 TTTTTCGATTTTATATATATATATATATATATATATATATATATATATAT 11110
 QY 307 TACTTCGTTTGA 320
 DB 11111 TTAATTTATTTTGA 11124

RESULT 11
 ABK33955
 ID ABK33955 standard; DNA; 14147 BP.
 AC ABK33955;
 XX
 DT 18-JUN-2002 (first entry)
 DE Human DNA for staging of Astrocytomas #20.
 XX
 KW Human: ds; astrocytoma; cytostatic; staging; cysteine methylation; Cpg;
 KW bisulphite; brain tissue; MALDI; ESI; election spray mass spectrometry;
 KW matrix assisted laser desorption/ionization mass spectrometry.
 XX
 OS Homo sapiens.
 XX
 PN WO200202808-A2.
 XX
 PD 10-JAN-2002.
 XX
 PE 02-JUL-2001; 2001WO-EP07538.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-171649/22.
 XX
 PT Novel chemically modified genomic DNA sequences, useful in the

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:09:40 ; Search time 12.5196 Seconds
(without alignments)
8769.450 Million cell updates/sec

Title: US-09-502-426a-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagtc.....tgtcaataaccacaacaag 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCRTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42.8	12.0	1493	1	US-08-340-820-24
2	42.8	12.0	1493	1	US-08-593-535-24
3	41.8	11.7	998	4	US-09-122-400B-5
4	41.1	11.5	2230	4	US-08-378-313-24
5	40.8	11.4	6243	2	US-09-056-075-1
6	40.8	11.4	19557	5	PCRT-US92-06300-1
7	40.6	11.3	1733	3	US-09-073-569-1
8	40.6	11.3	3095	6	5231168-1
9	40.4	11.3	10607	1	US-08-078-090-3
10	40.4	11.2	2755	3	US-08-749-522-2
11	39.4	11.0	289	1	US-08-341-568-3
12	39.4	11.0	289	1	US-08-911-020-3
13	39.4	11.0	19124	2	US-08-487-826B-13
14	39.4	11.0	1895	4	US-09-444-336-7
15	39.4	10.9	4285	4	US-09-410-464-1
16	39.4	10.9	9048	3	US-08-973-273-4
17	38.8	10.8	835	4	US-08-998-416-547
18	38.8	10.8	863	4	US-08-998-416-498
19	38.8	10.8	3138	1	US-07-867-106-4
20	38.2	10.7	2836	3	US-08-747-221B-24
21	38.2	10.7	2836	3	US-08-747-221B-26
22	38.2	10.7	2836	4	US-09-005-051-24
23	38.2	10.7	2836	4	US-09-005-051-26
24	38.2	10.7	5852	1	US-07-867-106-2
25	38.2	10.7	29604	3	US-08-781-891-207
26	38.2	10.6	552	4	US-09-134-001C-2606
27	38.2	10.6	701	4	US-08-998-416-701

28	38	10.6	2555	2	US-08-693-457-3	Sequence 3, Appl
29	38	10.6	2555	4	US-09-265-731-3	Sequence 3, Appl
30	38	10.6	8700	2	US-08-392-625-16	Sequence 16, Appl
31	38	10.6	8700	2	US-08-466-961A-16	Sequence 16, Appl
32	38	10.6	8700	2	US-08-645-193B-18	Sequence 18, Appl
33	37.8	10.6	1582	3	US-08-545-196B-10	Sequence 10, Appl
34	37.8	10.6	1582	3	US-08-545-196B-12	Sequence 12, Appl
35	37.6	10.5	2107	4	US-09-180-852-1	Sequence 1, Appl
36	37.6	10.5	2422	1	US-07-867-106-5	Sequence 5, Appl
37	37.6	10.5	3138	1	US-07-867-106-4	Sequence 4, Appl
38	37.6	10.5	5852	1	US-07-867-106-2	Sequence 2, Appl
39	37.6	10.5	6078	4	US-09-173-914-1	Sequence 1, Appl
40	37.4	10.4	477	4	US-08-887-534A-81	Sequence 81, Appl
41	37.4	10.4	2445	6	5215909-9	Patent No. 5215909
42	37.4	10.4	2674	4	US-09-817-180-1	Sequence 1, Appl
43	37.4	10.4	3761	4	US-08-890-865A-2	Sequence 2, Appl
44	37.4	10.4	10223	4	US-08-961-527-73	Sequence 73, Appl
45	37.2	10.4	144	1	US-08-702-344-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-340-820-24/c
Sequence 24, Application US/08340820
Patent No. 5512460
GENERAL INFORMATION:
APPLICANT: NABUO, Ken-ichi
APPLICANT: SEKO, Chisako
APPLICANT: KUROKAWA, Tsutomu
APPLICANT: KUNO, Tatsuya
TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
ADDRESSEE: CUSHMAN
STREET: 130 Water Street
City: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,820
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/835,713
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CONLIN, David G.
REGISTRATION NUMBER: 27026
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

	Query Match	11.5%;	Score 41;	DB 4;	Length 2230;
	Best Local Similarity	44.9%;	Pred. No. 2.1;		
	Matches 153; Conservative	0;	Mismatches 190;	Indels 0;	Gaps 0;
Oy	4 AAAAAAGATCAAAAGTATTTTATTCTCCTCTTTTTTTTGATAAATTTTAATCATTTT	63			
Dd	2222 AACATATATATATTT	2163			
Oy	64 TTTTGGCCCAATGATATATATAAAAATTGGANAATATATATTATGTGATATTCGTTTTTTA	123			
Dd	2162 TTTTATTTTAAATTAGCTAAGACAAATTAATGATGCTAAGACTATTTTTTTTTTTTTTTT	2103			
Oy	124 GTTCGGGTTTGGAAGAAGGTTTCGACTTTGCAAGTGGACATGTATATATATGGAG	183			
Dd	2102 TTAAACAATACCTTCATCTTTTGCATTCATTATATTAGCTAAGACAAATTAATGATGCTCAA	2043			
Oy	184 CTAGGTTGAGTCCTTGGACATTTGTAATGGATGTTGTGATATTAAGTCGACACTATT	243			
Dd	2042 CATTTTTTTTTTTTTTTTTTTTTTTTAAACATACCTTCACITTTGCATTCATTATATTAGCTA	1983			
Oy	244 AAACCTTAATGGGCTTCTCTATTAAGGCCCAATTATATTAAGTATTAACAAAAGTGCAAC	303			
Dd	1982 ACAAAATTAAGATGCTCAAGTACATTTTTTTTTTTTTTTTTTTTTTTCATAACATTAACCTTCACT	1923			
Oy	304 TTTTACTCGTGTTTTGATCCGAAGCAATTAACAATTTGTAATATC 348				
Dd	1922 TTTGCATTCATTAAATTAGCTAAGACAAATTAATGATGCTCAATAC 1878				

```

1      RESULT 5
2      US-09-056-075-1/c
3      : Sequence 1, Application US/09056075
4      : Patent No. 5953568
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: Johnson, Eric A.
8      : APPLICANT: Bradshaw, Marile
9      : APPLICANT: Rood, Julian
10     : TITLE OF INVENTION: Expression System for Clostridium
11     : TITLE OF INVENTION: Species
12     : NUMBER OF SEQUENCES: 2
13     :
14     : CORRESPONDENCE ADDRESSES:
15     : ADDRESSEE: Quarles & Brady
16     : STREET: 1 South Pinckney Street
17     : CITY: Madison
18     : STATE: WI
19     :
20     : COUNTRY: US
21     : ZIP: 53701-2113
22     :
23     : COMPUTER READABLE FORM:
24     : MEDIUM TYPE: Floppy disk
25     : COMPUTER: IBM PC compatible
26     : OPERATING SYSTEM: PC-DOS/MS-DOS
27     : SOFTWARE: PatentIn Release #1.0, Version #1.30
28     : CURRENT APPLICATION DATA:
29     : APPLICATION NUMBER: US/09/056.075
30     : FILING DATE:
31     :
32     : CLASSIFICATION:
33     : ATTORNEY/AGENT INFORMATION:
34     : NAME: Seay, Nicholas J.
35     : REGISTRATION NUMBER: 27386
36     : REFERENCE/DOCKET NUMBER: 960236.95238
37     : TELECOMMUNICATION INFORMATION:
38     : TELEPHONE: 608-251-5000
39     : TELEFAX: 608-251-9166
40     :
41     : INFORMATION FOR SEQ ID NO: 1:
42     : SEQUENCE CHARACTERISTICS:
43     : LENGTH: 6243 base pairs
44     : TYPE: nucleic acid
45     : STRANDEDNESS: double
46     : TOPOLOGY: linear
47     :
48     : MOLECULE TYPE: DNA (genomic)
49     : FEATURE:
50     : NAME/KEY: misc_feature
51     : LOCATION: 3770..4013
52     : OTHER INFORMATION: /note= "Rp4 origin of DNA transfer (oriT) from
53     : US-09-056-075-1 plasmid Rp4"

```

Query Match	11.4%	Score 40.8	DB 2	Length 6243
Best Local Similarity	47.3%	Pred. No.2.3		
Matches 155	Conservative	0	Mismatches 172	Indels 1
			Gaps	1
QY 4	AAAAAAGATCAAAAGTATTTTATCTCTCTCTTTTTTTTGGATAAATTTTAAATCATTT	63		
DB 1416	AAATAAAAAATATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	1357		
QY 64	TTTTTGCCCAATGATATATAAAAATTGGATA-AAATAATATTTGGATATTCGTTTTT	122		
DB 1356	TTTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	1297		
QY 123	ACTGGGGTTTGGAAAAAGGCTTTCGACTTCGAAAGCGACAGTATATGATTGGCA	182		
DB 1236	TTTATTTTTTTTATTTTTTATTTTTTATTTTTTATTTTTTATTTTTTATTTTTT	1237		
QY 183	GCTAGGTGAGCTTTGGACATTTGATTTGGANGTTGTGATTAATAGTGGACACTAT	242		
DB 1236	TTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	1177		
QY 243	TAAACCTTAAATGGGCTTCTCTAATGAAGGCCCAATATATATACGATTTAACAAAGTGACA	302		
DB 1176	ACGCTGTATTTATAGGCTACTTGTACCTGTCTTTTTTTTGGGAGGCTGTAAAGATATA	1117		
QY 303	CTTTTACTTTCGTTTTTGGATCCGAGCAA	330		

Db 1116 TTTTACTTGTAGTATGTCGAGGAA 1089

RESULT 6
PCT-US92-06300-1/c

Sequence 1, Application PC/TUS9206300

GENERAL INFORMATION:

APPLICANT: Hurwitz, David R

APPLICANT: Nathan, Margaret

APPLICANT: Shanl, Moshe

TITLE OF INVENTION: Transgenic Protein Production

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Rhone-Poulenc Rorer, Inc.

STREET: 500 Virginia Ave., Bldg. 3A

CITY: Ft. Washington

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19034

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/06300

FILING DATE: 1992/07/30

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Goodman, Rosanne

REGISTRATION NUMBER: 52,534

REFERENCE/DOCKET NUMBER: A0856-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 962-4107

TELEFAX: (215) 962-4130

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1957 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOHETICAL: NO

ANTI-SENSE: NO

PUBLICATION INFORMATION:

AUTHORS: Minshetti, P P

AUTHORS: Ruffner, D E

AUTHORS: Kuang, W.-J.

AUTHORS: Denison, O E

AUTHORS: Hawkins, J W

AUTHORS: Beattie, W G

AUTHORS: Dugalczyk, A

TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN GENE

TITLE: IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN Q11-22

JOURNAL: J. Biol. Chem.

VOLUME: 261

PAGES: 6747-6757

DATE: 1986

RELAVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 19002

PCT-US92-06300-1

Query Match 11.4%; Score 40.8; DB 5; Length 19557;

Best Local Similarity 52.3%; Pred. No. 2.3;

Matches 90; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Db 2 AAAAAAAAAAGTGAAGATTTTATCTCTCTTTTGTGATAATTTAAATCAT 61

Db 423 ATATCATATATATGATATATATTTCTTTTCTTTTCTTTTCTTTTCTTTT 364

Db 62 TTTTGGCCCAATGATATATAAATTTGATTAATAATATATATGATATCTTTT 121

Db 363 TTTTATATATCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 304

Db 122 TACTGCGGTTTGAGAAAGGTTTCGATTCGAAAGTGACGATATAT 173

Db 303 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 252

RESULT 7

US-09-073-569-1/c

Sequence 1, Application US/09073569

Patent No. 6084088

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Grossmann, Angelika

TITLE OF INVENTION: NOVEL TUMOR ANTIGENS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Zymogenetics, Inc.

STREET: 1201 Eastlake Avenue East

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98102

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/073,569

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sawilak, Deborah A

REGISTRATION NUMBER: 37,438

REFERENCE/DOCKET NUMBER: 97-14

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-442-6672

TELEFAX: 206-442-6678

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1733 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 34...1344

OTHER INFORMATION:

US-09-073-569-1

Query Match 11.3%; Score 40.6; DB 3; Length 1733;

Best Local Similarity 56.3%; Pred. No. 2.5;

Matches 76; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Db 20 TATTTTATCTCTCTCTTTTGTGATAATTTAAATCATTTTGGCCCAATGATA 79

Db 1715 TTTTATTTTGTGATAATTTAAATTTGATATGCTTTTGTGCGGTTTGAGAA 139

Db 1655 TTTTATTTTGTGATAATTTAAATTTGATATGCTTTTGTGCGGTTTGAGAA 1596

Db 140 AGGTTTGCATTTTC 154

Db 1595 AGGTCGTATATATC 1581

RESULT 8

5231168-1
; Patent No. 5231168
; APPLICANT: DIEBIELE, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; SEQ ID NO: 1:
; FILING DATE: 18-SEP-1989
; LENGTH: 3095
5231168-1
Query Match
Best Local Similarity 11.3%; Score 40.6; DB 6; Length 3095;
Matches 64; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 4 AAAAAAAAAAGATGATATTTATCTCTCTTTTGGATTAATTTAAATCATTT 63
DB 2724 AAGAAGAGAAAAATTTTATATATCATTTTTCATTAATAATTAATAATTA 2783
OY 64 TTTTGGCCATGATATATTAATAATTTGGATTAATAATATAT 106
DB 2784 TTAATATATAATATCGATTAATATTTATGATATATATATAT 2826
RESULT 9
US-08-078-090-3
; Sequence 3, Application US/08078090
; Patent No. 5739407
; GENERAL INFORMATION:
; APPLICANT: BERGSTROM, SVEN
; APPLICANT: HERNELL, OLLE
; APPLICANT: LOENNERDAL, BO
; APPLICANT: HJALMARSSON, KARIN
; APPLICANT: HANSSON, LENNART
; APPLICANT: TOERNELL, JAN
; APPLICANT: STROMQUIST, MATS
; TITLE OF INVENTION: HUMAN BETA-CASEIN PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 SEVENTH STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/078, 090
; FILING DATE: 19930618
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NO PCT/DK92/00236
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NO PCT/DK91/00233
; FILING DATE: 19-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28, 005
; REFERENCE/DOCKET NUMBER: BERGSTROM2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)628-5197
; TELEFAX: (202)737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

LENGTH: 10607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4804..4854, 5720..5746, 6726..6746, 6845
; ..6886, 7991..8521, 9440..9445)
US-08-078-090-3

Query Match
Best Local Similarity 11.3%; Score 40.4; DB 1; Length 10607;
Matches 74; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY 60 ATTTTGGCCCATGATATATAATAATTTGGATTAATATATATGATATTCGTT 119
DB 923 ATTATAGTATAGAAAGATGATATAAACTAGTATATATATATATATATAT 982
OY 120 TTTAGTGGGTTTGAGAAAGGTTTCGACTTTCGAAAGTGACGATATATATG 179
DB 983 ATTAGTATATATAATGATGATATATATATATATATATATATATATATAT 1042
OY 180 GGACCTAGCT 189
DB 1043 GTAGATTAGT 1052

RESULT 10
US-08-749-522-2
; Sequence 2, Application US/08749522
; Patent No. 6096950

; GENERAL INFORMATION:
; APPLICANT: John, Mallyakal
; TITLE OF INVENTION: FIBER-SPECIFIC PROMOTERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: U.S.A.
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,522
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 670513.90244
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 271-5709
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-749-522-2

Query Match
Best Local Similarity 11.2%; Score 40; DB 3; Length 2755;
Matches 171; Conservative 0; Mismatches 180; Indels 3; Gaps 2;
OY 6 AAAAAAAAAAGATATTTATCTCTCTTTTGGATTAATTTAAATCATTTT 65

US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 5993827

GENERAL INFORMATION:
 APPLICANT: Sim, Kim L.
 APPLICANT: Chitnis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhan
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 NUMBER OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 CORRESPONDENCE ADDRESS: 45

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10 SEP 1993
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Israelsten, Ned
REGISTRATION NUMBER: 29, 655
REFERENCE/DOCKET NUMBER: NH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
IS-08-487-826B-13

Best Local Similarity 46.7%; Pred. No. 4.4; Mismatches 161; Conservative 0; Mismatches 181; Indels 3; Gaps 1

QY	6	AAAAAGATGAAGCTATTTATTTATTCGCTCTTTTTTTTATGATTAATTTAAATCAATTTT	65
Db	15815	AAATTAATTTTTTTTATTATGATATATATTTTTTTTAACTTTTTTAAATTTTTTTT	15756
QY	66	TTTGCCCATGATATATTAATAAATTTGGATTAATAATATATATGATATTCGCTTTTTACT	125
Db	15755	TATTTATGATATATATTTTATTTTAAATATTTTTTTCTTTTTTGGTTTTATATA	15686
QY	126	TCGGTTTGAGAAAAGGGTTTGACTTTCGAAATGACATGATATATA--GATGGGA	182
Db	15695	TATATATTTTTTTTTTTTAAAGTTTTTTTTTTCTCTCTTGTTTTATTTTTTTTATA	15636

OY	183	GCTAGTGTGAGTCITTTGGACATTTGATTGGATGTTGTGTAATTAATAGTCGCACACTAT	242
Db	15635	ATCCTTTTTTTTTTTAATAAAATTTTTTTTTTAATTTTTTTTGANAAATCITTTCATTTT	15576
OY	243	TAAACCTTAATGGGCTTCTCAATGAAGCCCATTTAATACGATTATPACAAAAGTCACA	302
Db	15575	TTATTCATCAAAATTTAATATTTTAATTAATTTAATTTAATTTAAAAAATTTTTC	15516
OY	303	CCTTTACTCGTTTGTGATCCAGAACATTAACAATATGTCAAAT	347
Db	15515	CTTTTTTTTTTTTTTTTTTATTTAATAAATTTTTTTTTTANA	15471

US-09-444-336-7/c
; Sequence 7, Application US/09444336
; Patent No. 6410713

```

: GENERAL INFORMATION:
: APPLICANT: Guerrierro, Vincent
: APPLICANT: Raynes, Deborah A
: TITLE OF INVENTION: DNA Encoding Proteins That Inhibit Hsp70 Function
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/09/444,336
: CURRENT FILING DATE: 1999-11-19
: EARLIER APPLICATION NUMBER: US/09/444,336
: EARLIER FILING DATE: 1998-11-20
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 7
: LENGTH: 1895
: TYPE: DNA
: ORGANISM: Brachydanio rerio (zebra fish)
US-09-444-336-7

```

Matches	60; Conservative	0; Mismatches	35; Indels
Best Local Similarity	63.2%;	Pred. No. 5.2;	

QY 20 TATTTTATCTCTCTTTTTTTTGGAAATTTTAAATCTTTTTTTTGGCCCAATGATA 79
 Db 1862 TTGGTAAACACA 1803
 QY 80 TATTAATAATTTGGATTAATATATATATATATATTT 114
 Db 1802 GGTCGATGATGTATATTTTAAATATATCTCATTT 1768

; Sequence 1, Application US/09410464
; Patent No. 6395892

```

: GENERAL INFORMATION:
: APPLICANT: Strauss et al.
: TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
: TITLE OF INVENTION: poplar and other plant species.
: FILE REFERENCE: 53375
: CURRENT APPLICATION NUMBER: US/09/410,464
: CURRENT FILING DATE: 1999-10-01
: EARLIER APPLICATION NUMBER: 09/287,700
: EARLIER FILING DATE: 1999-04-06
: EARLIER APPLICATION NUMBER: 60/080,851
: EARLIER FILING DATE: 1998-04-06
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4285
: TYPE: DNA
: ORGANSIM: Populus balsamifera subsp. trichocarpa
: US-09-410-464-1

```

Best Local Similarity	55.68;	Pred. No. 5.2;	Longen 42.0;
Matches	75;	Conservative	0;
		Mismatches	60;
		Indels	0.
		Cons	0

```

OY 6 AAAAAGATGAAGATATTTTATCTCTCTTTTGTGATATTTAAATCATTTT 65
    ||||| | ||| | | | | | | | | | | | | | | | | | | | |
Db 1513 AAAAATATTTTGAATATTTTATTTTATATATAATTTTATAGATCATTTTA 1572
    ||||| | ||| | | | | | | | | | | | | | | | | | | | |
OY 66 TTGCCCATGATATATATAATTTGATATAATATATATGATATCGTTTCTAGT 125
    ||| | | | | | | | | | | | | | | | | | | | | | | |
Db 1573 ATACGTAAATATAAAAAATATTTTAAAAAAATTTTAAATATATTTTAAANA 1632
    ||| | | | | | | | | | | | | | | | | | | | | | | |
OY 126 TCGGGTTTGAGAAAA 140
    ||| | | | |
Db 1633 TAATATTTAAAAAA 1647
    ||| | | | |

```

Search completed: March 30, 2003, 09:01:56
 Job time : 33.5196 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:25:26 ; Search time 27.2667 Seconds

(without alignments)

11172.915 Million cell updates/sec

Title: US-09-502-426a-1_COPY_6111_6468

Perfect score: 358

Sequence: 1 aaaaaaaaaagatgaagf.....tgcataatcaacaacaag 358

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCFMS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	47.8	13.4	516 10 US-09-960-352-5785	Sequence 5785, Ap
2	44.5	12.6	189 9 US-10-102-806-401	Sequence 401, App
3	44.6	12.5	277 10 US-09-960-352-12673	Sequence 12673, A
4	44.6	12.5	681 10 US-09-822-830A-36	Sequence 26, App1
5	43.8	12.2	277 10 US-09-960-352-5094	Sequence 5094, Ap
6	43.8	12.2	446 10 US-09-960-352-3400	Sequence 3400, App
7	43.6	12.2	447 10 US-09-770-444-759	Sequence 2919, App
8	43.6	12.2	413 10 US-09-960-352-2919	Sequence 6528, Ap
9	43.2	12.1	414 10 US-10-028-072-345	Sequence 345, App
10	43.2	12.1	2933 9 US-10-121-043-345	Sequence 345, App
11	43.2	12.1	2933 9 US-10-123-904-345	Sequence 345, App
12	43.2	12.1	2933 9 US-10-140-470-345	Sequence 345, App
13	43.2	12.1	2933 9 US-10-175-746-345	Sequence 345, App
14	43.2	12.1	2933 9 US-10-176-921-345	Sequence 345, App
15	43.2	12.1	2933 9 US-10-176-921-345	Sequence 345, App
16	43.2	12.1	2933 9 US-10-137-865-345	Sequence 345, App
17	43.2	12.1	2933 9 US-10-140-474-345	Sequence 345, App
18	43.2	12.1	2933 9 US-10-142-431-345	Sequence 345, App
19	43.2	12.1	2933 9 US-10-142-431-345	Sequence 345, App

C 20	43.2	12.1	2933 9 US-10-143-114-345	Sequence 345, App
C 21	43.2	12.1	2933 9 US-10-140-002-345	Sequence 345, App
C 22	43.2	12.1	2933 9 US-10-142-419-345	Sequence 345, App
C 23	43.2	12.1	2933 9 US-10-123-262-345	Sequence 345, App
C 24	43.2	12.1	2933 9 US-10-142-423-345	Sequence 345, App
C 25	43.2	12.1	2933 9 US-10-121-050-345	Sequence 345, App
C 26	43.2	12.1	2933 9 US-10-141-755-345	Sequence 345, App
C 27	43.2	12.1	2933 9 US-09-828-366-20	Sequence 20, App
C 28	43.2	12.1	2933 9 US-09-880-107-3327	Sequence 3327, App
C 29	43.2	12.1	2933 9 US-09-754-853A-2	Sequence 3, App11
C 30	43.2	12.1	2933 9 US-09-754-853A-3	Sequence 3, App11
C 31	42.8	12.0	33513 9 US-09-960-352-12412	Sequence 12412, A
C 32	42.4	11.9	1344 9 US-09-938-842A-4296	Sequence 4296, Ap
C 33	42.4	11.8	253 10 US-09-878-574-13973	Sequence 13973, A
C 34	42.4	11.8	431 10 US-09-960-352-5538	Sequence 5558, Ap
C 35	42.4	11.8	202001 9 US-10-274-990-3	Sequence 3, App11
C 36	42.4	11.8	202001 10 US-09-734-674-3	Sequence 315, App
C 37	42.2	11.8	2045 9 US-10-013-477-2	Sequence 2, App11
C 38	42.2	11.8	2045 9 US-09-880-107-706	Sequence 706, App
C 39	41.8	11.7	329 10 US-09-960-352-1707	Sequence 1707, Ap
C 40	41.8	11.7	430 10 US-09-816-894-5	Sequence 5, App11
C 41	41.8	11.7	998 9 US-09-833-153-35	Sequence 35, App1
C 42	41.8	11.7	1820 9 US-09-960-352-4584	Sequence 4584, Ap
C 43	41.6	11.6	416 10 US-09-885-303A-11	Sequence 11, App1
C 44	41.6	11.6	3282 9 US-09-960-352-1243	Sequence 1243, Ap
C 45	41.4	11.6	291 10	

ALIGNMENTS

RESULT 1
US-09-960-352-5785
Sequence 5785, Application US/09960352
Patient No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Ningling
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OR INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT FILING DATE: US/09/960,352
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 5785
LENGTH: 516
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: unsure
LOCATION: (76), (90)
OTHER INFORMATION: unsure at all n locations
US-09-960-352-5785
Query Match
Best Local Similarity 13.4%; Score 47.8; DB 10; Length 516;
Matches 70; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

DB 410 TATTTTTCCTCTTTTTCGTAATTTTTCACATTTTTCGCCAATATA 79
OY 80 TATTAATTTTGGATTAATATATATTTTGGATTTTCGTTAGTT 126
DB 470 TATTTTTCCTCTTTTTCGTAATTTTTCACATTTTTCGCCAATATA 79
OY 80 TATTAATTTTGGATTAATATATATTTTGGATTTTCGTTAGTT 126
RESULT 2
US-10-102-806-401/C
Sequence 401, Application US/10102806

```

: Publication No. US20030054421A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA103P1C1
: CURRENT APPLICATION NUMBER: US/10/102,806
: CURRENT FILING DATE: 2002-03-22
: PRIOR APPLICATION NUMBER: 09/925,298
: PRIOR FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05861
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 846
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 401
: LENGTH: 189
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (11)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (162)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (165)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (166)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (187)
: OTHER INFORMATION: n equals a,t,g, or c
: US-102-806-401

Query Match          12.6%; Score 45; DB 9; Length 189;
Beat Local Similarity 63.3%; Pred. No. 2.1;
Matches 69; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 20 TATTTTATCTCTTCTTTTGTGATTAATTTTAAATCATTTTTTGGCCCAATGATA 79
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 158 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 99
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 80 TATAAAATTTGGATTAATAATATATATGATATCGTTTTTAACTTG 128
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98 TATTAATAATTAACATGATATAGCAGGTTTCAACAATGTCTGTACTTGG 50
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
US-09-960-352-12673/c
: Sequence 12673, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Mengling
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: FILE REFERENCE: 16511,006/37-21,(10298)C
: CURRENT APPLICATION NUMBER: US/09/960,352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 12673
: LENGTH: 277
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 54-ILIB3058-039-Q1-K1-F10

```

```

US-09-960-352-12673

Query Match      12.5%; Score 44.6; DB 10; Length 277;
Best Local Similarity 57.6%; Pred. No. 2.8;
Matches 80; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 2 AAAAAAAAAAGATGAAAGATATTTTATCTCTCTCTTTTGGATATTTAAATCAT 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 275 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTCOA 216

OY 62 TTTTTCGCCCATGATATATATATAAATTTGGATTAATATATATATTTGATATCGTTT 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 215 AATTTTTCATTTTAAATTTTTCCTTTTAAATTTTAAATTTTAAATTTT 156

OY 122 TAGTTGGGTTTGAGAAA 140
    | || ||| ||| |||
DB 155 TTTTTCATTTTAAAAAAA 137

RESULT 4
US-09-822-830A-26
; Sequence 26, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakari
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-26

Query Match      12.5%; Score 44.6; DB 10; Length 681;
Best Local Similarity 53.1%; Pred. No. 3.6;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 20 TATTTTATTCCTCTCTTTTGGTAATTTAAATCAATTTTGGCCATGATA 79
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 321 TTTTTCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 380

OY 80 TATATAAATTTGATATAATATATATATGATATCGTTTATAGTGGGTTGAGAAA 139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 381 TTTTTCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 440

OY 140 AGGGTTTCAGCTTCGAAAGTGGACGATGTATATAGATTTGGAGCTAGGTCTTT 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 441 ATAAGCTTATTTACATCAAGTAAATTAATATACATCAAAAGATGCAAAACGTTTATGTCATT 499

RESULT 5
US-09-960-352-5094
; Sequence 5094, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan

```

;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;; FILE REFERENCE: 16511.006/37-21(10298)C
;; CURRENT APPLICATION NUMBER: US/09/960,352
;; NUMBER OF SEQ ID NOS: 2001-09-24
;; SEQ ID NO 5094
;; LENGTH: 277
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (17),(37),(41)
;; OTHER INFORMATION: unsure at all n locations
US-09-960-352-5094

Query Match 12.2%; Score 43.8; DB 10; Length 277;
Best Local Similarity 57.8%; Pred. No. 4;
Matches 78; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGAGTATTTTCTCTTTTGGATTTTAAATCA 60
DB 143 AATAATTTATTTGACATTTGATTAATTTATTTTAAATTT 202
OY 61 TTTTGGCCCAATGATATATAAATTTGGATAATATTTGATTCGTTT 120
DB 203 TTTTATTTTATTTTATTTTATTTTAAATTTTATTTTAAATTTAT 262
OY 121 TTAATTCGGGTTTGA 135
DB 263 TTAATTTTATTTTGA 277

RESULT 6
US-09-960-352-3400
;; Sequence 3400, Application US/09960352
;; Patent No. US20020137139A1
;; GENERAL INFORMATION:
;; APPLICANT: Warren, Wesley C.
;; APPLICANT: Tao, Nengding
;; APPLICANT: Byatt, John C.
;; APPLICANT: Mathialagan, Nagappan
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;; FILE REFERENCE: 16511.006/37-21(10298)C
;; CURRENT APPLICATION NUMBER: US/09/960,352
;; CURRENT FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 15112
;; SEQ ID NO 3400
;; LENGTH: 446
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; OTHER INFORMATION: Clone ID: 15-LIB3058-052-Q1-K1-D11
US-09-960-352-3400

Query Match 12.2%; Score 43.8; DB 10; Length 446;
Best Local Similarity 49.4%; Pred. No. 4.6; Indels 0; Gaps 0;
Matches 114; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

OY 20 TATTTTATCTCTCTTTTGGATAATTTTAAATCATTTTGGCCCAATGATA 79
DB 143 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 202
OY 80 TATAAATTTGATAATATATTTGATATTTGTTTATTTAGTTCGGGTTTGAAA 139
DB 203 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 262
OY 140 AGGTTTCGACTTCGAAGTGAAGATGATATAGATTGGAGTGAAGTTCGTTTG 199
DB 263 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 322
OY 200 GACATTTGATTTGATTTGTTGATTTAGTGTGACACTATTAACCTT 250

DB 323 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 373

RESULT 7
US-09-770-444-759
;; Sequence 759, Application US/09770444
;; Patent No. US2002023280A1
;; GENERAL INFORMATION:
;; APPLICANT: Gorlach, Jorn
;; APPLICANT: An, Yong-Qiang
;; APPLICANT: Hamilton, Carol M.
;; APPLICANT: Price, Jennifer L.
;; APPLICANT: Raines, Tracy M.
;; APPLICANT: Rameaka, Joshua G.
;; APPLICANT: Page, Amy
;; APPLICANT: Mathew, Abraham V.
;; APPLICANT: Ledford, Brooke L.
;; APPLICANT: Woessner, Jeffrey P.
;; APPLICANT: Haas, William David
;; APPLICANT: Garcia, Carlos A.
;; APPLICANT: Krickler, Maja
;; APPLICANT: Slader, Ted
;; APPLICANT: Davis, Keith R.
;; APPLICANT: Allen, Keith
;; APPLICANT: Hoffman, Neil
;; APPLICANT: Hurban, Patrick
;; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
;; FILE REFERENCE: 2027 (PABA-016PRV)
;; CURRENT APPLICATION NUMBER: US/09/770,444
;; PRIOR FILING DATE: 2001-01-26
;; PRIOR APPLICATION NUMBER: 60/178,502
;; NUMBER OF SEQ ID NOS: 999
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 759
;; LENGTH: 447
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
US-09-770-444-759

Query Match 12.2%; Score 43.8; DB 10; Length 447;
Best Local Similarity 72.2%; Pred. No. 4.6;
Matches 57; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 20 TATTTTATCTCTCTTTTGGATAATTTTAAATCATTTTGGCCCAATGATA 79
DB 22 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 81
OY 80 TATAAATTTGATAATAT 98
DB 82 TCCAGAACTTTTATTAAT 100

RESULT 8
US-09-960-352-2919
;; Sequence 2919, Application US/09960352
;; Patent No. US20020137139A1
;; GENERAL INFORMATION:
;; APPLICANT: Warren, Wesley C.
;; APPLICANT: Byatt, John C.
;; APPLICANT: Mathialagan, Nagappan
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;; FILE REFERENCE: 16511.006/37-21(10298)C
;; CURRENT APPLICATION NUMBER: US/09/960,352
;; CURRENT FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 15112
;; SEQ ID NO 2919
;; LENGTH: 413

1 PRIOR APPLICATION NUMBER: 60/085323
2
3 PRIOR FILING DATE: 1998-05-13
4 PRIOR APPLICATION NUMBER: 60/085338
5 PRIOR FILING DATE: 1998-05-13
6 PRIOR APPLICATION NUMBER: 60/085339
7 PRIOR FILING DATE: 1998-05-13
8 PRIOR APPLICATION NUMBER: 60/085579
9 PRIOR FILING DATE: 1998-05-15
10 PRIOR APPLICATION NUMBER: 60/085657
11 PRIOR FILING DATE: 1998-05-15
12 PRIOR APPLICATION NUMBER: 60/085704
13 PRIOR FILING DATE: 1998-05-15
14 PRIOR APPLICATION NUMBER: 60/086414
15 PRIOR FILING DATE: 1998-05-22
16 PRIOR APPLICATION NUMBER: 60/086430
17 PRIOR FILING DATE: 1998-05-22
18 PRIOR APPLICATION NUMBER: 60/087106
19 PRIOR FILING DATE: 1998-05-28
20 PRIOR APPLICATION NUMBER: 60/088026
21 PRIOR FILING DATE: 1998-06-04
22 PRIOR APPLICATION NUMBER: 60/088730
23 PRIOR FILING DATE: 1998-06-10
24 PRIOR APPLICATION NUMBER: 60/088741
25 PRIOR FILING DATE: 1998-06-10
26 PRIOR APPLICATION NUMBER: 60/088810
27 PRIOR FILING DATE: 1998-06-10
28 PRIOR APPLICATION NUMBER: 60/088858
29 PRIOR FILING DATE: 19/98-06-11
30 PRIOR APPLICATION NUMBER: 60/089532
31 PRIOR FILING DATE: 1998-06-17
32 PRIOR APPLICATION NUMBER: 60/089599
33 PRIOR FILING DATE: 1998-06-17
34 PRIOR APPLICATION NUMBER: 60/089907
35 PRIOR FILING DATE: 1998-06-18
36 PRIOR APPLICATION NUMBER: 60/089947
37 PRIOR FILING DATE: 1998-06-19
38 PRIOR APPLICATION NUMBER: 60/090349
39 PRIOR FILING DATE: 1998-06-23
40 PRIOR APPLICATION NUMBER: 60/090429
41 PRIOR FILING DATE: 1998-06-24
42 PRIOR APPLICATION NUMBER: 60/090445
43 PRIOR FILING DATE: 1998-06-24
44 PRIOR APPLICATION NUMBER: 60/090538
45 PRIOR FILING DATE: 1998-06-24
46 PRIOR APPLICATION NUMBER: 60/090863
47 PRIOR FILING DATE: 1998-06-26
48 PRIOR APPLICATION NUMBER: 60/091360
49 PRIOR FILING DATE: 1998-07-01
50 PRIOR APPLICATION NUMBER: 60/091519
51 PRIOR FILING DATE: 1998-07-02
52 PRIOR APPLICATION NUMBER: 60/091982
53 PRIOR FILING DATE: 1998-07-07

RESULT 11
US-10-121-049-345/c
: Sequence 345, Application US/10121049
: Publication No. US2003002239A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Beresini, Maureen

ORGANISM: Homo Sapien
US-10-123-904-345
Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C17
CURRENT APPLICATION NUMBER: US/10/121,049
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-121-049-345

Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C17
CURRENT APPLICATION NUMBER: US/10/121,049
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-121-049-345

RESULT 12
US-10-123-904-345/C
Sequence 345, Application US/10123904
Publication No. US20030022326A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C14
CURRENT APPLICATION NUMBER: US/10/123,904
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA

ORGANISM: Homo Sapien
US-10-123-904-345

Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C17
CURRENT APPLICATION NUMBER: US/10/121,049
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-121-049-345

RESULT 13
US-10-140-470-345/C
Sequence 345, Application US/10140470
Publication No. US20030022331A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C16
CURRENT APPLICATION NUMBER: US/10/140,470
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-470-345

Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C16
CURRENT APPLICATION NUMBER: US/10/140,470
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-470-345

Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C16
CURRENT APPLICATION NUMBER: US/10/140,470
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-470-345

RESULT 14
US-10-175-746-345/C
Sequence 345, Application US/10175746
Publication No. US20030027270A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C16
CURRENT APPLICATION NUMBER: US/10/175,746
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:08:52 ; Search time 467.144 Seconds

(without alignments)
12411.569 Million cell updates/sec

Title: US-09-502-426a-1_COPY_6111_6468

Sequence: 358 1 aaaaaaaaaagatgaagt.....tgcaataccaacacaaag 358

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_hum:*
20: em_gss_pln:*
21: em_gss_vtc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137.4	38.4	176	17	AL762543 Arabidops
2	128.4	35.9	194	17	AL762544 Arabidops
3	121	33.8	423	10	AV787618 Arabidops
4	68	19.0	1101	17	CNS000T2 Arabidops
5	65.4	18.3	928	17	CNS00DKY Arabidops
6	60	16.8	905	17	CNS00KHX Arabidops

C	7	58.2	16.3	1027	17	CNS02T50	AT121733 Tetradon
C	8	58	16.2	1101	17	CNS016LI	AT106896 Drosophila
C	9	57.8	16.1	961	17	AZ691838	AT691838 ENTM156TF
C	10	57	15.9	1101	17	CNS00EJA	AL0692527 Drosophila
C	11	56.6	15.8	894	17	CNS018BG	AL109126 Drosophila
C	12	56.6	15.8	1147	17	CNS073CX	AL427351 Drosophila
C	13	56.4	15.8	1147	17	CNS0029N	AL097397 Drosophila
C	14	56.2	15.7	1101	17	CNS0021J	AL061936 Drosophila
C	15	56	15.6	949	17	CNS04AIH	AL281906 Tetradon
C	16	55.8	15.6	767	17	CNS00AIX	AL175966 Tetradon
C	17	55.6	15.5	1092	17	CNS00AOX	AL1419462 T3 end of
C	18	55	15.4	1007	17	CNS02OK7	AL065924 Drosophila
C	19	54.8	15.3	1101	17	CNS00IEB	AL059273 Drosophila
C	20	54.6	15.3	952	17	CNS006V8	AL065967 Drosophila
C	21	54.6	15.3	1002	17	CNS0185X	AL108927 Drosophila
C	22	54.6	15.3	1121	17	CNS01623	AL106197 Drosophila
C	23	54.4	15.2	423	17	CNS001W9	AL075520 Drosophila
C	24	54.2	15.1	866	17	CNS0216T	AL099876 Drosophila
C	25	54.2	15.1	1101	17	CNS00EVL	AL069706 Drosophila
C	26	54	15.1	952	17	CNS016RS	AL107122 Drosophila
C	27	54	15.1	993	9	AL561361	AL561361 AT561361
C	28	54	15.1	1200	17	CNS016CO	AL106578 Drosophila
C	29	53.6	15.0	1043	17	CNS0145P	AL103735 Drosophila
C	30	53.6	15.0	1094	17	CNS012R2	AL101513 Drosophila
C	31	53.6	15.0	1167	17	CNS07360	AL427102 clone BAO
C	32	53.4	14.9	973	17	CNS071LE	AL425064 clone BAO
C	33	53.4	14.9	974	17	AG043196	AG043196 Pan trogl
C	34	53.2	14.9	458	9	AL514085	AL514085 AT514085
C	35	53.2	14.9	994	17	CNS0058N	AL057169 Drosophila
C	36	53	14.8	573	17	CNS06LTX	AL040635 T7 end of
C	37	52.8	14.7	813	17	CNS03CTM	AL237857 Tetradon
C	38	52.8	14.7	637	17	CNS036CC	AL229845 Tetradon
C	39	52.6	14.7	1225	17	CNS0161D	AL106171 Drosophila
C	40	52.4	14.6	529	9	AL514657	AL514657 AT514657
C	41	52.4	14.6	1101	17	CNS0177R	AL107697 Drosophila
C	42	52.4	14.6	1201	17	CNS01038	AL098462 Drosophila
C	43	52.2	14.6	403	17	AO880113	AO880113 HS_4868_B
C	44	52.2	14.6	759	17	CNS060XV	AL411257 T7 end of
C	45	52.2	14.6	765	17	BH148938	BH148938 ENPT52TF

ALIGNMENTS

RESULT 1	AL762543	176 bp	DNA	linear	GSS 19-JUN-2002
LOCUS	AL762543				
DEFINITION	Arabidopsis thaliana T-DNA flanking sequence GK-025603-013722,				
ACCESSION	AL762543				
VERSION	AL762543.1	GI:21508636			
KEYWORDS	GSS.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H. and Weisshaar,B.				
TITLE	A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines				
JOURNAL	Unpublished				
REFERENCE	Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.				
AUTHORS	A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 176)				
AUTHORS	Li,Y., Strizhov,N., Rosso,M. and Weisshaar,B.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer				

COMMENT

Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone T3A5. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
plant genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

Location/Qualifiers

1..176
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-025G03-013722"
/note="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector PAC106. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"

BASE COUNT 60 a 33 c 27 g 56 t
ORIGIN

Query Match 38.4%; Score 137.4; DB 17; Length 176;
Best Local Similarity 99.3%; Pred. No. 1e-11;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 220 TTGATTATTAGTGCACACTTAACTTAAAGGCTCTCTAATAGCCCAATTATA 279

DB 1 TTGATTATTAGTGCACACTTAACTTAAAGGCTCTCTAATAGCCCAATTATA 60

OY 280 TTACGATTATACCAAGTGACAACTTTACTCGTTTGTATCGGAGCAATTAACAAT 339

DB 61 TTACGATTATACCAAGTGACAACTTTACTCGTTTGTATCGGAGCAATTAACAAT 120

OY 340 GTCAAAATACCAACACACAG 358

DB 121 GTCAAAATACCAACACACAG 139

RESULT 2

LOCUS

AL762544 194 bp DNA linear GSS 19-JUN-2002

DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-025G03-013757,
genomic survey sequence.

ACCESSION

AL762544

VERSION

AL762544.1 GI:21508641

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

1 Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H.
and Weisshaar, B.
A pipeline for automated high-throughput generation of FSTs
(flanking sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
Unpublished

JOURNAL

REFERENCE

AUTHORS

TITLE

Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
for flanking sequence tag based reverse genetics
Unpublished

JOURNAL

REFERENCE

AUTHORS

TITLE

Li, Y., Strizhov, N., Rosso, M. and Weisshaar, B.
Direct Submission
Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer

COMMENT

Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone T3A5. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
plant genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

Location/Qualifiers

1..194
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-025G03-013757"
/note="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector PAC106. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"

BASE COUNT 66 a 37 c 28 g 63 t
ORIGIN

Query Match 35.9%; Score 128.4; DB 17; Length 194;
Best Local Similarity 98.6%; Pred. No. 2.1e-10;
Matches 140; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 217 TTGTTATTATTAGTGCACACTTAACTTAAAGGCTCTCTAATAGCCCAATT 276

DB 27 TTGTTATTATTAGTGCACACTTAACTTAAAGGCTCTCTAATAGCCCAATT 85

OY 277 ATATTGATTATTAACAAGTGACAACTTTACTCGTTTGTATCGGAGCAATTAACA 336

DB 86 ATATTGATTATTAACAAGTGACAACTTTACTCGTTTGTATCGGAGCAATTAACA 145

OY 337 ATTGTCAAAATACCAACACACAG 358

DB 146 ATTGTCAAAATACCAACACACAG 167

RESULT 3

LOCUS

AV787618 423 bp mRNA linear EST 28-MAR-2002

DEFINITION AV787618 RAF16 Arabidopsis thaliana cDNA clone RAF106-76-F19 3',
mRNA sequence.

ACCESSION

AV787618

VERSION

AV787618.1 GI:19806408

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 Seki, M., Narusawa, M., Ishida, J., Kamuya, A., Satou, M., Nakaajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawaji, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). This clone is in a

JOURNAL

REFERENCE

AUTHORS

TITLE

BASE COUNT 262 a /note="end : 17"
ORIGIN 70 c 84 g 321 t 191 others

Query Match 18.3%; Score 65.4; DB 17; Length 928;
Best Local Similarity 31.0%; Pred. No. 0.27;
Matches 99; Conservative 83; Mismatches 137; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGAGTATTTTATCTCTCTTTTGTGATATTTAAATCA 60
DB 524 AAAAAAAAAATTTTWTWTTTAAATTTTWTWTTTAAATTTTAAATTT 583
OY 61 TTTTTCGCCAATGATATATAATTTGATATATATATATGATATCTTTT 120
DB 584 ATTTTWTATTAATAAATAAATAAATAAATAAATAAATAAATAAATA 643
OY 121 TTAGTCGGGTTTGAGAAAGGTTTCGACTTTCGAAAGTATATAGATTGG 180
DB 644 WTATTTTATATATAAATAAATAAATAAATAAATAAATAAATAAATA 703
OY 181 GAGCTAGGTTGAGTCTTGGACATTTGATTTGATTTAGTGTGACACT 240
DB 704 TTATTTTATTTTATTTTATTTATTTATTTATTTATTTATTTATTT 763
OY 241 ATTTAACTTAATGGCTTCTATAGCCCAATTTATTTAGATTTATACAAATGAC 300
DB 764 WTTTWTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 823
OY 301 AACTTTTACTCTTTTGTG 319
DB 824 KAKKKKKKKKKKKKKKK 842

RESULT 6

CNS00KHX 905 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACR17N06 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL077798.1 GI:4957174
VERSION GSS.
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 905)
AUTHORS Genoscope.

COMMENT Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)

COMMENT

The BACP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BACP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammose in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BACP from the
isogenic strain Y2; cn bw sp, the same strain used for the BACP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source Location/Qualifiers
1..905
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR17N06"

BASE COUNT 388 a /clone_lib="RPCI-98"
ORIGIN 82 c 77 g 194 t 164 others

Query Match 16.8%; Score 60; DB 17; Length 905;
Best Local Similarity 35.9%; Pred. No. 1.8;
Matches 127; Conservative 59; Mismatches 168; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGAGTATTTTATCTCTCTTTTGTGATATTTAAATCA 60
DB 882 AAAAAAAAAATTTTWTWTTTAAATTTTWTWTTTAAATTTTAAATTT 823
OY 61 TTTTTCGCCAATGATATATAATTTGATATATATATGATATCTTTT 120
DB 822 WTTTWTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 763
OY 121 TTAGTCGGGTTTGAGAAAGGTTTCGACTTTCGAAAGTATATAGATTGG 180
DB 762 TTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 703
OY 181 GAGCTAGGTTGAGTCTTGGACATTTGATTTGATTTAGTGTGACACT 240
DB 702 TTTTTCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 643
OY 241 ATTTAACTTAATGGCTTCTATAGCCCAATTTATTTAGATTTATACAAATGAC 300
DB 642 TTTTTCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 583
OY 301 AACTTTTACTCTTTTGTG 354
DB 582 AAAAAAAAAATTTTWTWTTTAAATTTTWTWTTTAAATTTTAAATTT 529

RESULT 7

CNS02T50 1027 bp DNA linear GSS 15-MAY-2000
LOCUS Tetradon nlgroviidids genome survey sequence T7 end of clone
DEFINITION 163M16 of library G from Tetradon nlgroviidids, genomic survey
sequence.

ACCESSION AL212733.1 GI:7871552
VERSION GSS; genome survey sequence.
KEYWORDS Tetradon nlgroviidids.
SOURCE Tetradon nlgroviidids.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetradontiformes;
Tetradontidae; Tetradon.

REFERENCE 1 (bases 1 to 1027)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Mincker, P., Brothier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.

COMMENT Human gene number estimate provided by genome wide analysis using
Tetradon nlgroviidids DNA sequence

REFERENCE 2 (bases 1 to 1027)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.

COMMENT Characterization and repeat analysis of the compact genome of the
freshwater puterfish Tetradon nlgroviidids
Unpublished
3 (bases 1 to 1027)

COMMENT Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
genome. For more information project of the Tetradon nlgroviidids
<http://www.genoscope.cns.fr/Tetradon>.

FEATURES
source Location/Qualifiers
1..1027

..... 120

Best Local Similarity
10.1%; Score 57.8; DB 17; Length 961;
64.7%; Pred. No. 3.8;

[illegible]

